Digitized by Arya Samaj Foundation Chennai and eGangotri

Digitized by Arya Samaj Foundation Chennai and eGangotri

J. Phil.

Vol. 84
1987

G. K. U.
Hardwar

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

110147

Digitized by Arya Samaj Foundation Chennai and eGangotri

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

पुस्तकालय
गुरुकुल कांगड़ी विश्वविद्यालय
हरिद्वार
7/3/87

THE JOURNAL OF PHILOSOPHY

5:11
L.

VOLUME LXXXIV, NUMBER 1
JANUARY 1987

page

5 *Indeterminacy of Translation Again* — W. V. Quine

11 *Explaining Cultural Divergence: A Wittgensteinian Paradox* — Paul Seabright

28 *A Philosophical Letter of Alfred Tarski* — Morton White

BOOK REVIEWS

33 Saunders MacLane: *Mathematics: Form and Function* — Colin McLarty

37 Chad Hansen: *Language and Logic in Ancient China* — Michael R. Martin

42 Anthony Kenny: *The Legacy of Wittgenstein* — Malcolm Budd

45 Patrick Suppes: *Probabilistic Metaphysics* — Henry E. Kyburg, Jr.

49 Myles Brand: *Intending and Acting: Toward a Naturalized Action Theory* — Peter Slezak

54 NOTES AND NEWS

56 MEMO TO AUTHORS

57 JOHNSONIAN PRIZE IN PHILOSOPHY

110147

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Sidney Morgenbesser, Charles D. Parsons, James J. Walsh. *Consulting Editor:* James T. Higginbotham. *Managing Editor:* Leigh S. Cauman. *Editorial & Business offices:* 720 Philosophy Hall, Columbia University, New York City 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $20.00; institutions, $25.00; students and the retired and unemployed, $15.00; plus $5 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *Acting President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Stephen S. Fenichell; Barbara Gimbel; André Schiffrin.

NOTE TO CONTRIBUTORS

As a monthly periodical, THE JOURNAL OF PHILOSOPHY encourages philosophical exchange and progress through prompt publication of articles of current interest. The editors now feel that they can ensure publication within six to twelve months of acceptance. They continue to be particularly interested in papers that explore the borderline between philosophy and the special disciplines, but are unlikely to accept articles primarily expository or historical in character. For the Comments and Criticism section they accept rejoinders but prefer those designed to advance the discussion rather than just to set the record straight. If and only if articles are held five months or more and then rejected, letters of rejection include a brief comment concerning the editors' reasons for rejection.

Attention is drawn to the memo to authors, republished on page 56 of this issue, and to the statement of copyright policy, which will appear in February; copies are available from the JOURNAL office.

Contributors are supplied free of charge with ten copies of the JOURNAL issues in which their articles appear, reviewers and authors of abstracts with two copies. Offprints may be ordered from the printer on a form supplied by the printer and enclosed with galley proof; they will be sent out within six weeks of publication.

© Copyright 1987 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy*, 720 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Panini Kanya Maha Vidyalaya Collection. Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIV, NUMBER 2
FEBRUARY 1987

715787
पुस्तकालय
गुरुकुल कांगड़ी विश्वविद्यालय
हरिद्वार

page

59 *Human Beings* — Mark Johnston
84 *The Dead* — Palle Yourgrau
102 *Moral Theory and Moral Alienation* — Adrian M. S. Piper

119 MEMO ON COPYRIGHT
120 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Sidney Morgenbesser, Charles D. Parsons, James J. Walsh. *Consulting Editor:* James T. Higginbotham. *Managing Editor:* Leigh S. Cauman. *Editorial & Business offices:* 720 Philosophy Hall, Columbia University, New York City 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $20.00; institutions, $25.00; students and the retired and unemployed, $15.00; plus $5 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *Acting President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Stephen S. Fenichell; Barbara Gimbel; André Schiffrin.

THEORY AND INTERPRETATION IN THE VISUAL ARTS

A Summer Institute for College and University Professors
Sponsored by Hobart and William Smith Colleges
With the Cooperation of the University of Rochester
Supported by the National Endowment for the Humanities
At Hobart and William Smith Colleges, Geneva, New York
July 12-August 21, 1987

For further information and application materials, please write or call: Tamar March, Associate Provost, Hobart and William Smith Colleges, Geneva, New York 14456-3397 (telephone 315: 789-5500).

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition, however, that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1987 by the Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices.

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIV, NUMBER 3
MARCH 1987

पुस्तकालय
गुरुकुल कांगड़ी विश्वविद्यालय
हरिद्वार

page

123 *Indeterminacy, Empiricism, and the First Person*
John R. Searle

147 *The Epistemology of First-person Reference*
Carol A. Rovane

168 *A Note on Extension, Intension, and Truth*
Anil Gupta and Nuel Belnap

175 NEW BOOKS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Sidney Morgenbesser, Charles D. Parsons, James J. Walsh. *Consulting Editor:* James T. Higginbotham. *Managing Editor:* Leigh S. Cauman. *Editorial & Business offices:* 720 Philosophy Hall, Columbia University, New York City 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $20.00; institutions, $25.00; students and the retired and unemployed, $15.00; plus $5 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *Acting President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Stephen S. Fenichell; Barbara Gimbel; André Schiffrin.

Journal of Philosophical Logic

Editor in Chief
RICHMOND H. THOMASON
University of Pittsburgh, U.S.A.

The *Journal of Philosphical Logic* is the only journal specializing in the philosophical studies utilizing formal methods or dealing with topics in logical theory.

Subscription Information ISSN 0022–3611
1986, Volume 15 (4 issues)
Institutional rate: Dfl. 211.00/US$ 76.00 per volume incl. postage/handling
Private rate: Dfl. 70.00/US$ 25.00 per volume incl. postage/handling
Private subscriptions should be sent direct to the publishers.

D. REIDEL PUBLISHING COMPANY

A Member of the Kluwer Academic Publishers Group

P.O. Box 17, 3300 AA Dordrecht, The Netherlands
190 Old Derby Street, Hingham MA 02043, U.S.A.
Falcon House, Queen Square, Lancaster LA1 1RN, U.K.

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition, however, that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

© Copyright 1987 by the Journal of Philosophy, Inc.

ISSN 0022-362X

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

पुस्तकालय
गुरुकुल कांगड़ी विश्वविद्यालय
हरिद्वार
27/7/07

VOLUME LXXXIV, NUMBER 4
APRIL 1987

page

181 *Functions* John Bigelow and Robert Pargetter

197 *Content by Courtesy* Lynne Rudder Baker

COMMENTS AND CRITICISM

214 *Atemporal Duration: A Reply to Fitzgerald* Eleonore Stump and Norman Kretzmann

220 *Does "Fitness" Fit the Facts? A Reply to Williams and Rosenberg's Rejoinder* Elliott Sober

BOOK REVIEWS

224 Alexander Rosenberg: *The Structure of Biological Science* Robert N. Brandon

227 Sydney Shoemaker: *Identity, Cause, and Mind* Colin McGinn

233 NOTES AND NEWS

234 NEW BOOKS: ANTHOLOGIES

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Aryà Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Sidney Morgenbesser, Charles D. Parsons, James J. Walsh. *Consulting Editor:* James T. Higginbotham. *Managing Editor:* Leigh S. Cauman. *Editorial & Business offices:* 720 Philosophy Hall, Columbia University, New York City 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $20.00; institutions, $25.00; students and the retired and unemployed, $15.00; plus $5 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *Acting President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Stephen S. Fenichell; Barbara Gimbel; André Schiffrin.

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition, however, that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1987 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *The Journal of Philosophy*, 720 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

पुस्तकालय
गुरुकुल कांगड़ी विश्वविद्यालय
हरिद्वार
17/18

VOLUME LXXXIV, NUMBER 5
MAY 1987

page

237 *Kantian Ethics and Supererogation* — Marcia Baron

263 *Moral Realisms and Moral Dilemmas* — Walter Sinnott-Armstrong

COMMENTS AND CRITICISM

277 *Scheffler on the Independence of Agent-centered Preogatives from Agent-centered Restrictions* — Larry A. Alexander

BOOK REVIEWS

284 Ronald Dworkin: *A Matter of Principle* and *Law's Empire* — Kent Greenawalt

292 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Sidney Morgenbesser, Charles D. Parsons, James J. Walsh. *Consulting Editor:* James T. Higginbotham. *Managing Editor:* Leigh S. Cauman. *Editorial & Business offices:* 720 Philosophy Hall, Columbia University, New York City 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $20.00; institutions, $25.00; students and the retired and unemployed, $15.00; plus $5 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *Acting President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Stephen S. Fenichell; Barbara Gimbel; André Schiffrin.

The Journal of Aesthetics and Art Criticism

VOLUME XLV Fall 1986 No. 1

Editorial
Art and Its Spectators—*David Carrier*
In Defense of Pictorial Mimesis—*David Blinder*
A Kuhnian Metatheory for Aesthetics—*Daniel Shaw*
Gadamer's Hermeneutics and the Uses of Forgery—*Ian Mackenzie*
An Interpretation of Sublimity, Ugliness, and Formlessness in Kant's Aesthetic Theory—*Theodore Gracyk*
Art and Interaction—*Noël Carroll*
Hume's Key and Aesthetic Rationality—*Steven Sverdlik*
Paul Valery's Modernist Aesthetic Object—*Steven Cassedy*
Afterwords
Book Reviews

Editor: John Fisher, Temple University, Philadelphia, Pa. 19122
Annual Subscription: $20.00 to individuals

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition, however, that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

© Copyright 1987 by the Journal of Philosophy, Inc.

ISSN 0022-362X

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

पुस्तकालय
गुरुकुल कांगड़ी विश्वविद्यालय
हरिद्वार
21910

VOLUME LXXXIV, NUMBER 6
JUNE 1987

page

293 *Identity, Essence, and Indiscernibility* — Stephen Yablo

315 *Van Fraassen on Explanation* — Philip Kitcher and Wesley Salmon

COMMENTS AND CRITICISM

331 *Mainsprings of Metaphor* — Catherine Z. Elgin and Israel Scheffler

BOOK REVIEWS

336 Richard Wollheim: *The Thread of Life* — Norton Batkin

344 ANNOUNCEMENTS

345 NOTES AND NEWS

ublished by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

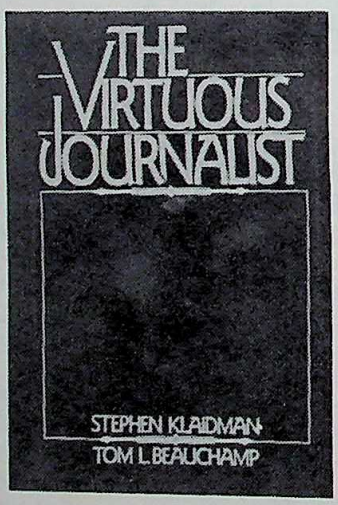
THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Sidney Morgenbesser, Charles D. Parsons, James J. Walsh. *Consulting Editor:* James T. Higginbotham. *Managing Editor:* Leigh S. Cauman. *Editorial & Business offices:* 720 Philosophy Hall, Columbia University, New York City 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $20.00; institutions, $25.00; students and the retired and unemployed, $15.00; plus $5 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *Acting President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Stephen S. Fenichell; Barbara Gimbel; André Schiffrin.



$19.95
OXFORD UNIVERSITY PRESS
Dept. NW, 200 Madison Avenue,
New York, NY 10016

"Solid, detailed comment on the ethics of stories still fresh in our minds."—*Columbia Journalism Review*

Reporting and editing the news poses a daily array of tough moral problems. In this book, a seasoned journalist and a respected philosopher combine their insights to produce a penetrating guide to journalistic ethics. Drawing on a wealth of real-life examples—from the CBS/Westmoreland case to coverage of the AIDS epidemic—Klaidman and Beauchamp treat such issues as notions of morality and fairness, standards of objectivity and accuracy, and maintaining public trust.

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1987 by the Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices.

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

पुस्तकालय
गुरुकुल कांगड़ी विश्वविद्यालय
हरिद्वार
21917

VOLUME LXXXIV, NUMBER 7
JULY 1987

page

349 *Red and Right* — Alan H. Goldman

COMMENTS AND CRITICISM

363 *Epistemic Tit for Tat* — Michel J. Blais

BOOK REVIEWS

376 Gilbert Harman: *Change in View: Principles of Reasoning* — Isaac Levi

384 Barry Stroud: *The Significance of Philosophical Scepticism* — Charles Larmore

392 J. F. Ross: *Portraying Analogy* — Josef Stern

398 ANNOUNCEMENTS

398 NEW BOOKS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

THE JOURNAL OF PHILOSOPHY
FOUNDED BY FREDERICK J.E. WOODBRIDGE AND WENDELL T. BUSH
Digitized by Arya Samaj Foundation Chennai and eGangotri

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Sidney Morgenbesser, Charles D. Parsons, Gisela Striker, James J. Walsh. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Charles Larmore. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $25.00; institutions, $40.00; students and the retired and unemployed, $20.00; plus $6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; Stephen S. Fenichell; and André Schiffrin.

MIND

Editor: Simon Blackburn, Pembroke College, Oxford

Mind is the best known British philosophical journal, and one of the most widely read and respected journals in professional philosophy. Founded in 1876, for over a century it has represented the leading philosophical ideas of its time.

Forthcoming and recent papers include *George Boolos* on Frege, *Michael Tye* on the Subjective Character of Experience, *André Gallois* on Rigid Designation and the Contingency of Identity, *Jennifer Hornsby* on Non Indicatives in Truth Theories, and *Susan Hurley* on Supervenience.

Rates for Volume 95, 1986: UK £10 N. America $22.50 Elsewhere £13.50 Institutions: UK £16 N. America $32.50 Elsewhere £19

XA

OXFORD JOURNALS
Journals Subscription Department, OUP, Walton Street, Oxford OX2 6DP, UK

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to other kinds of copying.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

© Copyright 1987 by the Journal of Philosophy, Inc.

ISSN 0022-362X

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

19191
गुरुकुल कांगड़ी विश्वविद्यालय
हरिद्वार

VOLUME LXXXIV, NUMBER 8
AUGUST 1987

page

405 *Utilitarianism and Expected Utility* John Broome

423 *Compatibilism and the Argument from Unavoidability* Thomas P. Flint

441 *Deciding to Believe without Self-deception* J. Thomas Cook

BOOK REVIEWS

447 Norbert Hornstein: *Logic as Grammar* Scott Soames

456 NEW BOOKS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Sidney Morgenbesser, Charles D. Parsons, Gisela Striker, James J. Walsh. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Charles Larmore. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

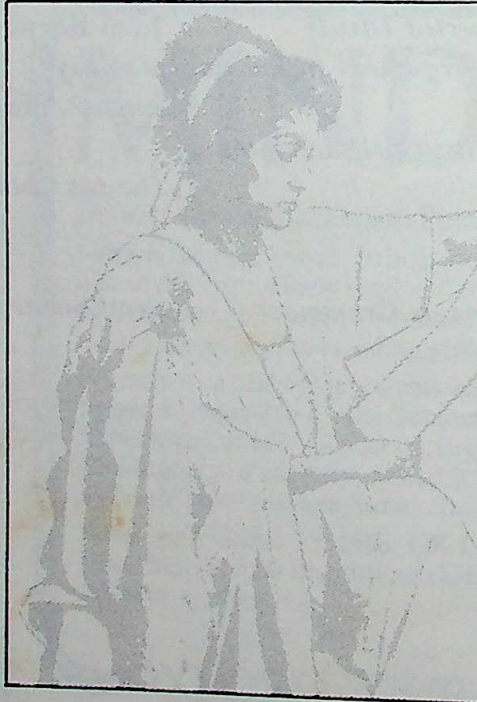
All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $25.00; institutions, $40.00; students and the retired and unemployed, $20.00; plus $6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; Stephen S. Fenichell; and André Schiffrin.

Hypatia

A JOURNAL OF

Feminist Philosophy



Subscriptions:

Individuals $20.00,
Institutions $40.00,

Foreign Orders add
$10 airmail. $5 surface

Editor: Margaret A. Simons
Hypatia
Southern Illinois University
at Edwardsville
Edwardsville, Illinois 62026-1437

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1987 by the Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

ISSN 0022-362X

... paid at New York, NY, and additional mailing offices.

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

3/11/07
पुस्तकालय
गुरुकुल कांगड़ी विश्वविद्यालय
हरिद्वार

VOLUME LXXXIV, NUMBER 9
SEPTEMBER 1987

page

465 *Freedom and Value* — Paul Benson

487 *A First Look at the Pornography/Civil Rights Ordinance: Could Pornography Be the Subordination of Women?* — Melinda Vadas

BOOK REVIEWS

512 M. J. Inwood: *Hegel* — Rüdiger Bittner

516 Hans Blumenberg: *Lebenszeit und Weltzeit* — Karsten Harries

520 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Sidney Morgenbesser, Charles D. Parsons, Gisela Striker, James J. Walsh. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Charles Larmore. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $25.00; institutions, $40.00; students and the retired and unemployed, $20.00; plus $6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; Stephen S. Fenichell; and André Schiffrin.

Journal of Indian Philosophy

Editor
B. K. Matilal, *All Souls College, Oxford, UK*

Indian philosophy has attracted a small audience in the West for many years, but it is only recently that Western philosophers have shown any general inclination to join it. The *Journal of Indian Philosophy* encourages this inclination and has been able to stimulate creative activities among orientalists and philosophers along with all the various combinations that these two classes can form. Contributions to the journal are bound by the limits of rational inquiry and avoid questions that lie in the fields of speculative sociology and para psychology. In a very general sense, the method is analytical and comparative, aiming at a rigorous precision in the translation of terms and statements. Space is devoted to the works of philosophers of the past as well as to the creative researches of contemporary scholars on such philosophic problems as were addressed by past philosophers.

Subscription Information **ISSN** 0022–1791
1987, Vol. 15 (4 issues)
Institutional rate: Dfl. 236.00/US$ 99.00 incl. postage/handling
Private rate: Dfl. 86.00/US$36.00 incl. postage/handling
Private subscriptions should be sent direct to the publishers

Back Volume(s) Available
Volumes 1–14 (1970–1986)
Volume 3 may be ordered from:
Swets & Zeitlinger BV, P.O. Box 810, 2160 SZ LISSE, The Netherlands

Price per Volume
excl. postage
Dfl. 195.00/US$68.00

Kluwer academic publishers group

P.O. Box 989
3300 AZ Dordrecht
The Netherlands

101 Philip Drive
Norwell,
MA 02061, U.S.A.

Falcon House
Queen Square
Lancaster, LA1 1RN, U.K.

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1987 by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Publication office: 709 Philosophy, Columbia U, NYC 10027.

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

पुस्तकालय
गुरुकुल कांगड़ी विश्वविद्यालय
हरिद्वार

VOLUME LXXXIV, NUMBER 10
OCTOBER 1987

Special Issue, Eighty-fourth Annual Meeting
American Philosophical Association, Eastern Division

page

Symposium: Husserl and Frege

521 *Husserl and Frege* — Robert Sokolowski
528 *Husserl and Frege* — Ronald McIntyre
535 *Frege and Husserl on Sense* — Donn Welton

Symposium: Epistemology and Philosophy of Mind

537 *Cognitive Science and Metaphysics* — Alvin I. Goldman
544 *Epistemology in the Age of Neuroscience* — Patricia Smith Churchland
553 *The Boundary between Philosophy and Cognitive Science* — George Bealer

Symposium: Liberties, Wrongs, and Representation

556 *Liberties, Wrongs, and Representation* — Jay Drydyk
565 *Political Representation and Economic Liberty* — Martin J. De Nys

Symposium: Conceptions of Causality in Early Modern Philosophy

567 *How God Causes Motion: Descartes, Divine Sustenance, and Occasionalism* — Daniel Garber
580 *Divine Impulse, Voluntary Motion, and Intelligible Causal Models* — Eileen O'Neill

582 NOTES AND NEWS

585 Program of the Meetings
597 List of Program Participants
602 Appendix: Programs of Group Meetings
613 List of Group Participants

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

THE JOURNAL OF PHILOSOPHY
FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH
Digitized by Arya Samaj Foundation Chennai and eGangotri

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Sidney Morgenbesser, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Charles Larmore. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $25.00; institutions, $40.00; students and the retired and unemployed, $20.00; plus $6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; Stephen S. Fenichell; and André Schiffrin.

IDEALISTIC STUDIES

An International Journal of Philosophy
Published by Clark University Press
Editor: Walter E. Wright

Issued three times a year, *Idealistic Studies* is a journal for studies of philosophical idealism. Both historical and contemporary statements of idealistic argumentation are welcomed, as are also critical studies of idealism.

Editorial Advisory Board: John P. Anton (Emory), Peter A. Bertocci (Boston), J.N. Findlay (Boston), Errol E. Harris (Northwestern), Reinhard Lauth (Munich), H.D. Lewis (London), Nicholas Rescher (Pittsburgh), Paul Ricoeur (Paris), George A. Schrader, Jr. (Yale), Charles E. Scott (Vanderbilt).

Yearly subscription rates for individuals, $12.00; libraries and institutions, $25.00. Single copies: $4.00.

All communication regarding manuscripts, subscriptions, and publication should be sent to the Editor, Department of Philosophy, Clark University, Worcester, Massachusetts 01610.

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

© Copyright 1987 by the Journal of Philosophy, Inc. ISSN 0022-362X

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIV, NUMBER 11
NOVEMBER 1987

19/3/19

Special Issue, Eighty-fourth Annual Meeting
American Philosophical Association, Eastern Division

page

Symposium: Feminist Perspectives on Individual Choice
619 *Personal Autonomy and the Paradox of Feminine Socialization* — Diana T. Meyers
628 *Autonomy and Respect* — Kathryn Pyne Addelson

Symposium: Mind Matters
630 *Mind Matters* — Ernest Le Pore and Barry Loewer
642 *Making Mind Matter More* — Jerry Fodor

Symposium: Relevant Alternatives and Demon Skepticism
643 *Relevant Alternatives and Demon Skepticism* — Bredo C. Johnsen
653 *Relevant Alternatives, Presuppositions, and Skepticism* — Jonathan E. Adler

Symposium: Public Policy and the Allocation of Scarce Medical Resources
655 *Public Policy and the Allocation of Scarce Medical Resources* — Richard L. Barber
664 *Justice and the Dissemination of "Big-ticket" Technologies* — Norman Daniels

Symposium: Aristotle's *Metaphysics*
666 *Aristotle's Solution to the Problem of Sensible Substance* — Edward Halper
673 *Hylomorphism in Aristotle* — Charlotte Witt
679 *Problems of Aristotle's Concept of Form* — Sarah Broadie

Symposium: Enlightenment and Rationality
682 *The Discourse of Modernity: Hegel and Habermas* — Fred Dallmayr
692 *Enlightenment and Rationality* — Axel Honneth
699 *Enlightenment and Reason* — Tom Rockmore

702 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Sidney Morgenbesser, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Charles Larmore. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $25.00; institutions, $40.00; students and the retired and unemployed, $20.00; plus $6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; Stephen S. Fenichell; and André Schiffrin.

REVUE INTERNATIONALE DE PHILOSOPHIE

Editor: Michel MEYER

143, av. A. Buyl, 1050 Brussels, Belgium

Each number is devoted to a particular movement, a particular philosopher, or a particular problem.

We publish 3 issues annually. Articles are written in English, French, German, or Italian.

Bacon

Philosophical aspects of literary theory

Our last issue was devoted to *Common Sense*

(articles by T. Sprigge, F. Jacques, M. G. Singer, W. Kluback, D. Schulthess, G. P. Pappas, F. Armengaud and A. R. White)

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1987 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy*, 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

19/3/88
पुस्तकालय
गुरुकुल कांगड़ी विश्वविद्यालय
हरिद्वार

VOLUME LXXXIV, NUMBER 12
DECEMBER 1987

page

707 *Serious Philosophy and Freedom of Spirit* — Ernest Sosa

727 *The Well-rounded Life* — Thomas Hurka

BOOK REVIEWS

747 Isaac Levi: *Decisions and Revisions: Philosophical Essays on Knowledge and Value* — Peter Gärdenfors

752 Jürgen Habermas: *Der philosophische Diskurs der Moderne: Zwölf Vorlesungen* — Seyla Benhabib

758 INDEX TO VOLUME LXXXIV

762 ANNOUNCEMENTS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Sidney Morgenbesser, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Charles Larmore. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $25.00; institutions, $40.00; students and the retired and unemployed, $20.00; plus $6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; Stephen S. Fenichell; and André Schiffrin.

U.S. Postal Service

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

Required by 39 U.S.C. 3685)

1A. TITLE OF PUBLICATION	1B. PUBLICATION NO.	2. DATE OF FILING
The Journal of Philosophy	0 0 2 2 3 6 2 X	9/23/87

9. FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 423.12 DMM only)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*

(1) [X] HAS NOT CHANGED DURING PRECEDING 12 MONTHS

(2) [ ] HAS CHANGED DURING PRECEDING 12 MONTHS

(If changed, publisher must submit explanation of change with this statement.)

10. EXTENT AND NATURE OF CIRCULATION *(See instructions on reverse side)*	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NO. COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A. TOTAL NO. COPIES *(Net Press Run)*	4845	4737
B. PAID AND/OR REQUESTED CIRCULATION 1. Sales through dealers and carriers, street vendors and counter sales	----	----
2. Mail Subscription *(Paid and/or requested)*	3961	3908
C. TOTAL PAID AND/OR REQUESTED CIRCULATION *(Sum of 10B1 and 10B2)*	3961	3908
D. FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	174	178
E. TOTAL DISTRIBUTION *(Sum of C and D)*	4135	4086
F. COPIES NOT DISTRIBUTED 1. Office use, left over, unaccounted, spoiled after printing	710	651
2. Return from News Agents	----	----
G. TOTAL *(Sum of E, F1 and 2 – should equal net press run shown in A)*	4845	4737

11. I certify that the statements made by me above are correct and complete

SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER

Michael Kelly, Managing Editor

PS Form 3526, July 1984 *(See instruction on reverse)*

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

Copyright-Free. In Public Domain. Gurukul Kangri Collection, Haridwar

© Copyright 1987 by the Journal of Philosophy, Inc.

ISSN 0022-362X

Digitized by Arya Samaj Foundation Chennai and eGangotri

पुस्तकालय
गुरुकुल कांगड़ी विश्वविद्यालय
हरिद्वार
7/3/82

THE JOURNAL OF PHILOSOPHY

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY was founded in 1904 as THE JOURNAL OF PHILOSOPHY, PSYCHOLOGY AND SCIENTIFIC METHODS by Frederick J. E. Woodbridge and J. McKeen Cattell. In 1906 Wendell T. Bush became associated with the JOURNAL as co-editor. In 1923, the JOURNAL was incorporated in the State of New York under its present name. Since its founding it has been published from Columbia University and printed by the Lancaster Press in Lancaster, Pennsylvania.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIV
JANUARY–DECEMBER 1987

Edited by
Bernard Berofsky, Arthur C. Danto
Hide Ishiguro, Isaac Levi, Mary Mothersill
Sidney Morgenbesser, Charles D. Parsons
James J. Walsh

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition, however, that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© COPYRIGHT 1987 BY THE JOURNAL OF PHILOSOPHY, INC.

PRINTED BY LANCASTER PRESS, INC., LANCASTER, PA.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

ISSN 0022-362X

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIV, NO. 1, JANUARY 1987

INDETERMINACY OF TRANSLATION AGAIN

OVER twenty-five years my *Word and Object* has sparked frequent criticism of my thesis of indeterminacy of translation and occasional evidence of misunderstanding, to which I have responded only sporadically and in scattered places. Burton Dreben has now drawn me back into fruitful discussions of the subject and urged me to undertake a succinct over-all clarification.

Critics have said that the thesis is a consequence of my behaviorism. Some have said that it is a *reductio ad absurdum* of my behaviorism. I disagree with this second point, but I agree with the first. I hold further that the behaviorist approach is mandatory. In psychology one may or may not be a behaviorist, but in linguistics one has no choice. Each of us learns his language by observing other people's verbal behavior and having his own faltering verbal behavior observed and reinforced or corrected by others. We depend strictly on overt behavior in observable situations. As long as our command of our language fits all external checkpoints, where our utterance or our reaction to someone's utterance can be appraised in the light of some shared situation, so long all is well. Our mental life between checkpoints is indifferent to our rating as a master of the language.

There is nothing in linguistic meaning, then, beyond what is to be gleaned from overt behavior in observable circumstances. In order to exhibit these limitations, I propounded the thought experiment of radical translation. The "source language," as the jargon has it, is Jungle; the "target language" is English. Jungle is inaccessible through any known languages as way stations, so our only data are native utterances and their concurrent observable circumstances. It is a meager basis, but the native speaker himself has had no other.

Our linguist would construct his manual of translation by conjec-

0022-362X/87/8401/0005$00.50 © 1987 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tural extrapolation of such data, but the confirmations would be sparse. Usually the concurrent publicly observable situation does not enable us to predict what a speaker even of our own language will say, for utterances commonly bear little relevance to the circumstances outwardly observable at the time; there are ongoing projects and unshared past experiences. It is only thus, indeed, that language serves any useful communicative purpose; predicted utterances convey no news.

There are sentences, however, that do hinge pretty strictly on the concurrent publicly observable situation: sentences like 'It's raining' or 'That's a rabbit', which I call *observation sentences*. Jungle sentences of this sort are our linguist's entering wedge. He tentatively associates a native's utterance with the observed concurrent situation, hoping that it might be simply an observation sentence linked to that situation. To check this he takes the initiative, when the situation recurs, and volunteers the sentence himself for the native's assent or dissent.

This expedient of query and assent or dissent embodies, in miniature, the advantage of an experimental science such as physics over a purely observational science such as astronomy. To apply it the linguist must be able to recognize, if only conjecturally, the signs of assent and dissent in Jungle society. If he is wrong in guessing those signs, his further research will languish and he will try again. But there is a good deal to go on in identifying those signs. For one thing, a speaker will assent to an utterance in any circumstance in which he would volunteer it.

Our linguist then goes on tentatively identifying and translating observation sentences. Some of them are perhaps compounded of others of them, in ways hinting of our logical particles 'and', 'or', 'but', 'not'. By collating the situations that command the natives' assent to the compounds with the situations that command assent to the components, and similarly for dissent, the linguist may get a plausible line on such connectives.

Unlike observation sentences, most utterances resist correlation with concurrent stimulations that the linguist can share. Taking the initiative, he may volunteer and query such a sentence for assent or dissent in various situations, but no correlation with concurrent stimulation is forthcoming. What next?

He can keep a record of these unconstrued sentences and dissect them. Some of the segments will have occurred also in the already construed observation sentences. He will treat them as words, and try pairing them off with English expressions in ways suggested by

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

those observation sentences. Such are what I have called analytical hypotheses. There is guesswork here, and more extravagant guesswork to follow. The linguist will turn to the unconstrued, nonobservational sentences in which these same words occurred, and he will project conjectural interpretations of some of those sentences on the strength of these sporadic fragments. He will accumulate a tentative Jungle vocabulary, with English translations, and a tentative apparatus of grammatical constructions. Recursion then sets in, determining tentative translations of a potential infinity of sentences. Our linguist keeps testing his system for its efficacy in dealing with natives, and he goes on tinkering with it and guessing again. The routine of query and assent that had been his standby in construing observation sentences continues to be invaluable at these higher and more conjectural levels.

Clearly the task is formidable and the freedom for conjecture is enormous. Radical translation is avoided in practice by finding someone who can interpret the language, however haltingly, into a somewhat familiar one. But it is only radical translation that exposes the poverty of ultimate data for the identification of meanings.

Let us consider, then, what constraints our radical translator can bring to bear to help guide his conjectures. Continuity is helpful: successive utterances may be expected to have some bearing on one another. When several such have been tentatively interpreted, moreover, their interrelation itself may suggest the translation of a linking word that will be helpful in spotting similar connections elsewhere.

The translator will depend early and late on psychological conjectures as to what the native is likely to believe. This policy already governed his translations of observation sentences. It will continue to operate beyond the observational level, deterring him from translating a native assertion into too glaring a falsehood. He will favor translations that ascribe beliefs to the native that stand to reason or are consonant with the native's observed way of life. But he will not cultivate these values at the cost of unduly complicating the structure to be imparted to the native's grammar and semantics, for this again would be bad psychology; the language must have been simple enough for acquisition by the natives, whose minds, failing evidence to the contrary, are presumed to be pretty much like our own. Practical psychology is what sustains our radical translator all along the way, and the method of his psychology is empathy: he imagines himself in the native's situation as best he can.

Our radical translator would put his developing manual of translation continually to use, and go on revising it in the light of his

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

successes and failures of communication. And wherein do these successes and failures consist, or how are they to be recognized? Successful negotiation with natives is taken as evidence that the manual is progressing well. Smooth conversation is further favorable evidence. Reactions of astonishment or bewilderment on a native's part, or seemingly irrelevant responses, tend to suggest that the manual has gone wrong.

We readily imagine the translator's ups and downs. Perhaps he has tentatively translated two native sentences into English ones that are akin to each other in some semantic way, and he finds this same kinship reflected in a native's use of the two native sentences. This encourages him in his pair of tentative translations. So he goes on blithely supposing that he is communicating, only to be caught up short. This may persuade him that his pair of translations was wrong after all. He wonders how far back, in the smooth-flowing antecedent conversation, he got off the beam.

Considerations of the sort we have been surveying are all that the radical translator has to go on. This is not because the meanings of sentences are elusive or inscrutable; it is because there is nothing to them, beyond what these fumbling procedures can come up with. Nor is there hope even of codifying these procedures and then *defining* what counts as translation by citing the procedures; for the procedures involve weighing incommensurable values. How much grotesqueness may we allow to the native's beliefs, for instance, in order to avoid how much grotesqueness in his grammar or semantics?

These reflections leave us little reason to expect that two radical translators, working independently on Jungle, would come out with manuals acceptable to both. Their manuals might be indistinguishable in terms of any native behavior that they gave reason to expect, and yet each manual might prescribe some translations that the other translator would reject. Such is the thesis of indeterminacy of translation.

My much-discussed example *gavagai*[1] illustrates the indeterminacy of translation only of terms, not of sentences. When seen rather as an observation sentence, *Gavagai* is keyed directly to determinate stimulus situations that are open to empirical investigation and afford the firmest of checkpoints. Likewise my example of Japanese classifiers[2] is a matter only of terms and not of sentences. More

[1] *Word and Object* (Cambridge, Mass.: MIT Press, 1960), pp. 29–45.
[2] *Ontological Relativity and Other Essays* (New York: Columbia, 1969), pp. 35–38.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

extravagant examples, having to do still with terms, are afforded by proxy functions.[3] But my thesis of indeterminacy of translation applies first and foremost to sentences, holophrastically conceived; and this I despair of documenting. Radical translation is a near miracle, and it is not going to be done twice to the same language. But surely, when we reflect on the limits of possible data for radical translation, the indeterminacy is not to be doubted.

The point of my thought experiment in radical translations was philosophical: a critique of the uncritical notion of meanings and, therewith, of introspective semantics. I was concerned to expose its empirical limits. A sentence has a meaning, people thought, and another sentence is its translation if it has the same meaning. This, we see, will not do.

The critique of meaning leveled by my thesis of indeterminacy of translation is meant to clear away misconceptions, but the result is not nihilism. Translation remains, and is indispensable. Indeterminacy means not that there is no acceptable translation, but that there are many. A good manual of translation fits all checkpoints of verbal behavior, and what does not surface at any checkpoint can do no harm.

Nor, in scouting the old notion of meanings of words and sentences, do I repudiate semantics. There is much useful work, done and to be done, regarding the manner and circumstances of the use of words. Lexicography is its conspicuous manifestation, and there is much scope also for refinements of semantic theory. But I would not seek a scientific rehabilitation of something like the old notion of separate and distinct meanings; that notion is better seen as a stumbling block cleared away. In later years indeed it has been more of a stumbling block for philosophers than for scientific linguists, who, understandably, have simply found it not technically useful.

Some of my readers have had difficulty in seeing how the thesis of indeterminacy of translation is anything but a special case of the thesis that natural science is underdetermined by all possible observation. This, in turn, is suggested by Pierre Duhem's perception that when we revise a theory in the light of a recalcitrant observation we are free to choose which component sentences of the theory to revoke.

The indeterminacy of translation differs from the underdetermination of science in that there is only the natives' verbal behavior for the manuals of translation to be right or wrong about; no claims are

[3] *Theories and Things* (Cambridge, Mass.: Harvard, 1981), pp. 19–22.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

laid regarding hidden neural mechanisms. If translators disagree on the translation of a Jungle sentence but no behavior on the part of the Jungle people could bear on the disagreement, then there is simply no fact of the matter. In the case of natural science, on the other hand, there is a fact of the matter, even if all possible observations are insufficient to reveal it uniquely. The facts of nature outrun our theories as well as all possible observations, whereas the traditional semantics outruns the facts of language.

In thus contrasting the underdetermination of natural science with the indeterminacy of translation I have taken a realistic view of nature, which indeed I hold. But I have elsewhere drawn the contrast without the realism, in the following way. Natural science, we again assume, is underdetermined by all possible observation. However, suppose that we have settled for one of the many over-all theories of nature that fit all possible observation. Translation remains indeterminate, even relative to the chosen theory of nature. Thus the indeterminacy of translation is an indeterminacy additional to the underdetermination of nature.

In closing I want to add an unrelated remark on indeterminacy, in the light of recurring evidence of misinterpretation. My indeterminacy applies only to translation, and in no way to grammaticality. I have represented grammaticality as subject to rounding out and rounding off at its limits, but apart from that I see grammaticality, unlike translation, as adequately determined by behavioral dispositions. Structurally unlike codifications of grammaticality can indeed be equivalent in output, but that goes without saying. Incidentally, semantics likewise continues to be, I repeat, a vital domain of inquiry. What I have challenged is just an ill-conceived notion within traditional semantics, namely, sameness of meaning.

W. V. QUINE

Harvard University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

EXPLAINING CULTURAL DIVERGENCE: A WITTGENSTEINIAN PARADOX*

> The problems are solved not by giving new information but by arranging what we have always known. Philosophy is a battle against the bewitchment of our intelligence by means of language.
>
> *Philosophical Investigations*†

THE part played in Wittgenstein's later philosophy of language by "forms of life"—that is, shared "practices," "customs," or "institutions"—has long been recognized to be crucial. Variously interpreted, forms of life underwrite the Wittgensteinian conception of rule following and the related notion of meaning in language, by deliberate contrast with the mentalistic entities to which explanatory appeal is made by the theories Wittgenstein was seeking to subvert. The part they play is so crucial, in fact, that there has also been a tendency among some commentators to treat forms of life as though they were fundamental categories, resistant to further analysis, either conceptual or empirical. But one difficulty with thus interpreting Wittgenstein's approach to language is that it leaves ambiguous the status of that subfamily of language-games called "social science," whose apparent role is to reflect on—and perhaps to explain—those very forms of life themselves and to account for the divergence in respect of them between different societies and cultures.

The purpose of this paper is to argue that this is a serious ambiguity. In particular I want to claim that two influential ways of interpreting Wittgenstein's approach to our use of language (or of what we might call "natural science" in a rather catholic sense of the term[1]) have the unacknowledged but bizarre consequence of ruling out completely the possibility of *social* science in any interesting sense. By 'social science' I mean the discipline that consists in the attempt to explain the existence and operation of those normative patterns of behavior which constitute shared human practices, cus-

* I am grateful to John Broome and Christopher Peacocke for helpful and detailed comments. I have also benefited from reading a draft of John McDowell's "Wittgenstein on Following a Rule" (see fn 7 below), which has been more influential on my thinking than my few references to it would suggest.

† Lugwig Wittgenstein, *Philosophical Investigations,* translated by G. E. M. Anscombe (Oxford: Blackwell, 1972), hereafter referred to as PI; §109.

[1] Corresponding perhaps to what Quine calls "the totality of our so-called knowledge or beliefs"; see "Two Dogmas of Empiricism," in *From a Logical Point of View* (New York: Harper & Row, 1961), p. 42.

0022-362X/87/8401/0011$01.70 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

toms, and institutions.[2] Although the impossibility of such a discipline is a conclusion that might appear to be supported by a superficial reading of Wittgenstein himself, I shall largely take it for granted that it is an untenable conclusion, both in fact and as interpretation; to attribute a view so startlingly opposed to common sense to someone who urged that philosophy "leaves everything as it is" (PI §124) would take some doing. Rather this consequence will be used to illuminate what I take to be wrong with these interpretations, and it will then be discussed what the role of social science would be in a cleaned-up version that preserved their important insights while avoiding their excesses.

What form, then, does the paradox take? Its general form is this: for Wittgenstein, forms of life are what underlie our rule following when "I have exhausted the justifications, [when] I have reached bedrock, and my spade is turned. Then I am inclined to say: 'This is simply what I do' " (PI §217). He says elsewhere: "But these explanations must after all sometime come to an end"[3] and forms of life are what remain. "When I obey a rule, I do not choose. I obey the rule *blindly*" (PI §219).

These views provide the basis for the first of the assumptions that generate our paradox:

(1) Forms of life are what remain *after* explanations have come to an end.

This assumption[4] I shall take to be both true and consistent with Wittgenstein's views. It is tempting to think that it entails the following:

(1a) Forms of life cannot be explained.

But this together with

(2) Social science is an attempt to explain forms of life.

entails

(3) Social science cannot succeed.

We can distinguish several possible reactions to this conclusion. One is to make a Sisyphean virtue of despair, and to claim that social

[2] The sense of the word 'explain' here—in particular, whether the explanation must be causal in character—is deliberately left open at this point.

[3] *On Certainty*, translated by Denis Paul and G. E. M. Anscombe (Oxford: Blackwell, 1969), §34.

[4] It is arguably really a definition, albeit one with substantive effect because of its being linked with assumption 2. See below.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

science remains a worth-while (or compelling) endeavor in spite of its being doomed to failure. Another is to give it up altogether. No further space will be devoted to considering these possibilities. A third consists in denying assumption 2, which might itself be done in one of several ways. First, it might be argued that not all human customs, practices, and institutions constitute forms of life in the requisite sense, and those that do not are the appropriate domain of social science. It is hard to think of conclusive arguments that might settle this claim either way, but there must count against it the consideration that it endows forms of life with a mystery and an air of technical wizardry which are quite at odds with Wittgenstein's continued emphasis on finding solutions in the everyday and the familiar—"don't think, but look!" (PI §66). It leaves mysterious, that is, the features that allegedly distinguish those practices which constitute forms of life from the rest. An alternative way to deny assumption 2 is to claim that social science is an attempt to explain forms of life, but one that uses a different conception of explanation from that invoked in assumption 1. It might, for instance, confine itself to a simple listing of the various features of human social practices. Though it is not impossible that a plausible argument along these lines might be constructed, it seems unlikely in view of the widespread conception of the social sciences as enabling in some form an enhanced understanding of the *meaning* of human customs and institutions, which is precisely the form of explanation to which assumption 1 makes reference.

The only remaining escape from the paradox is to deny that (1a) follows from (1). But to do so is not entirely straightforward. One tempting way is to use the analogy of Neurath's boat, which must be rebuilt plank by plank while still afloat. "Assumption 1," it might be said, "doesn't mean that our forms of life are not open to examination, merely that in examining some we must appeal to others. Wittgenstein is just warning us against the folly of imagining that we can step outside our practices, outside language, altogether."

Now although, as we shall see, this reply undoubtedly states an important truth, it is not very illuminating. For the tension between assumption 1 and our natural inclination to say "Why do these forms of life exist? Why do people behave in such a fashion?" appears to go deeper than this. At several points Wittgenstein appears to be saying that our appeal to forms of life precludes our explaining *them* at all. In particular, causal histories do not do any explanatory work. Consider the following:

> If the formation of concepts can be explained by facts of nature, should we not be interested, not in grammar, but rather in that nature which is

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

> the basis of grammar? . . . But our interest does not fall back upon these possible causes of the formation of concepts; we are not doing natural science, nor yet natural history (PI II xii).[5]

> But I don't mean that what I do now (in grasping a sense) determines the future use causally and as a matter of experience, but that in a queer way, the use itself is in some sense present! Really the only thing wrong with what you say is the expression 'in a queer way'. The rest is all right (PI §195).

> In our failure to understand the use of a word we take it as the expression of a queer *process* (PI §196).

> You must bear in mind that the language-game is so to say something unpredictable. I mean it is not based on grounds. It is not reasonable (or unreasonable). It is there—like our life (*On Certainty* §559).

> Let me ask this: what has the expression of a rule—say a sign-post—got to do with my actions? What sort of connexion is there here?—Well, perhaps this one: I have been trained to react to this sign in a particular way, and now I do so react to it. But that is only to give a causal connexion; to tell how it has come about that we now go by the sign post; not what this going-by-the-sign really consists in. On the contrary; I have further indicated that a person goes by a sign-post only in so far as there exists a regular use of signposts, a custom (PI §198).

What Wittgenstein seems to be saying here is not that the establishment of causal connexions is impossible, but that it is empty, that it does not tell us what a custom or form of life consists in or how it can underwrite our following of a rule. It does not tell us what makes physical movements and the production of sounds count as *behavior*. In short, there seems no room for social science to explain the meaning of human behavior at all. And in one passage, where he writes "What has to be accepted, the given, is—so one could say—forms of life" (PI, II, xi, §226). Wittgenstein appears to make their inexplicability *definitive* of forms of life.

This, then, is the paradox in its general form. Bernard Williams touches upon it in his essay "Wittgenstein and Idealism",[6] where he remarks rightly on its relationship to the old problem "that there is the gravest difficulty . . . in both positing the independent existence of culturally distinct groups with different world-views, and also of holding that any access we have to them is inescapably and non-trivially conditioned by our own world-view." However, Williams seems

[5] It should be noted that Wittgenstein uses the terms 'natural science' and 'natural history' to refer to what I prefer to call "social science," leaving 'natural science' for a broader, more Quinean use.

[6] In his *Moral Luck* (New York: Cambridge, 1981).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

in the end happy to attribute such a position to Wittgenstein, even though he tries to suggest that it is less alarming than it looks at first. I want to suggest that these passages are consistent with Wittgenstein's rejecting the paradox. But before I back up this claim I want to show how assumption 1a—and hence the paradox—is entailed by two specific interpretations of Wittgenstein's remarks on rule following. The first is what I shall call the "objectivist" interpretation, held by, for example, Jonathan Lear,[7] and the second is the "skeptical" interpretation advanced by Saul Kripke.[8]

To take Lear first, it should be stressed that it is not his objectivism as such which is the problem. It is the way the objectivism of his interpretation is supported. Lear asks "Why must philosophy be done by pointing? Because we cannot step outside our form of life and discuss it like some *objet trouvé*." He invites us to consider the alternative answers to the following questions:

1. What does 7 + 5 equal?
   (a) 12
   (b) Anything at all, just as long as everyone is so minded.
2. What follows from *P* and *If P then Q*?
   (a) *Q*
   (b) Anything at all, just as long as everyone is so minded.

"To each of these questions," as he rightly says, "(a) gives the correct answer." But why? According to Lear, this is because

> The possibility of there being persons who are minded in any way at all is the possibility of their being minded as we are. Our problem is that being minded as we are is not one possibility that we can explore among others . . . there is no genuine possibility of having fundamentally different routes of interest and perceptions of salience, for that is the spurious possibility of becoming other-minded. The illusion of possibility is engendered by considering our form of life as one among others (386).

[7] In "Leaving the World Alone," this JOURNAL, LXXIX, 6 (June 1982): 382–403. I do not for a moment suggest that this is true of all interpretations of Wittgenstein as "objectivist." For example, everything I say in this paper is consistent with John McDowell's remarks in "Wittgenstein on Following a Rule," *Synthese*, LVIII, 3 (March 1984): 325–363. I cannot make up my mind, however, about Barry Stroud, "Wittgenstein and Logical Necessity," in George Pitcher, ed., *Wittgenstein: The Philosophical Investigations* (London, Macmillan, 1966). Stroud makes several remarks which seem to imply that Wittgenstein embraces the paradoxical conclusion, but at other points his interpretation seems much more subtle.

[8] In *Wittgenstein on Rules and Private Language* (Oxford: Blackwell, 1982). In what follows I shall use the term 'Kripke's view' to refer to the view of Wittgenstein as interpreted by Kripke, without intending any attribution of views to Kripke himself.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Read straight, this seems to saddle Wittgenstein with the paradox with a vengeance. Forms of life are not open to being considered as "one among others"—and *a fortiori* the question "Why this form of life and not another?" cannot be answered. Yet Lear's argument obviously enshrines an important insight. What might this be? The reason we find it natural to say that only (a) is the correct answer, I suggest, is that, for Wittgenstein, the fact of agreement is neither a justification for the truth of our judgments nor a part of what is meant by them. Answer 2b appears to imply either

(2b′) To say that Q is the answer just is to say that everyone agrees that Q is the answer.

or

(2b″) It is everyone's agreeing that Q is the answer that justifies our giving Q as the answer.

It is not necessary to show the untenability of either of these views, for both are clearly repudiated by Wittgenstein. Of (2b′) he writes that "the proposition 'Human beings believe that twice two is four' and 'twice two is four' do not mean the same" (PI, II, xi, p. 226), and (2b″) is rejected in these words:

> So you are saying that human agreement decides what is true and what is false?—It is what human beings say that is true and false; and they agree in the *language* they use. This is not agreement in opinions but in form of life (PI II, xi, p. 241).

Yet it is not necessary to go all the way with Lear. For the claim that:

> [T]he fact of agreement is neither a justification for what we say nor a part of what we mean by it.

does not entail assumption 1a:

> We cannot consider or give any explanation of our form of life as one among others.

Indeed, it is hard to see that there should be any prima facie connection between them, and to appeal to the latter in support of the correctness of Lear's answer (a) is simply redundant when the former provides much simpler and better support.

The attractiveness of assumption 1a is no doubt due in part to the support it appears to receive from such remarks of Wittgenstein as the following:

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

> I have not said *why* mathematicians do not quarrel, but only that they do not (PI, II, xi, p. 226).

There is a view, however, that puts a less skeptical construction on this remark. It is that there *are* ways of explaining our forms of life—of considering them as possibilities among others—which do not themselves constitute *justifications* of our ways of carrying on. In other words, there are ways of accounting causally and nonemptily for forms of life being thus and not otherwise, in which the fact of our meaning this and not that can consist, but which we may not legitimately appeal to as *justification* for our carrying on thus and not otherwise. This view respects Wittgenstein's insistence that agreement *underlies* rather than is expressed by or appealed to by our judgments. It merely adds that it is perfectly legitimate to ask "*why* mathematicians do not quarrel," only not in the same breath as when we do mathematics. Only this view leaves open the possibility of a social science.

This claim will be clarified later, after an examination of Kripke's skeptical view. Like Lear, Kripke believes that agreement must be seen as underlying rather than constituting meaning, though he puts the point in a nihilistic way. He stresses the importance of assertability rather than truth conditions in Wittgenstein's later view of what establishes the correctness of linguistic assertions. He then writes, of intersubjective agreement, that:

> What follows from these assertability conditions is *not* that the answer everyone gives to an addition problem is, by definition, the correct one, but rather the platitude that if everyone agrees upon a certain answer, then no one will feel justified in calling the answer wrong (§112).

The reason it is put thus is that Kripke has argued, by an application of Goodman's Paradox, that "there can be no fact as to what I mean by 'plus', or any other word at any time" (p. 21). At least, this is the view he attributes to Wittgenstein. In the face of it, his "sceptical solution" (that the fact of agreement removes the urge to disagree) is a weak and jejune concession to our inclination to believe, as John McDowell puts it, "that to learn the meaning of a word is to acquire an understanding that obliges us subsequently . . . to judge and speak in certain determinate ways" (325). According to Kripke, there is no fact in which this understanding consists, and the possibility of social science is cut away from below. The reason we cannot "explain" our practices is that there is nothing to them, no fact in which they consist. There is no fact that constitutes our understanding being of such a kind that it will in the future lead us to behave and

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

make judgments in a certain way. We just do things, and that is all there is to it. Our forms of life just are the fact that on the occasions in question we have done this, this, and this. There is, so to speak, no structure to them.

This view would, incidentally, be one way of explaining *why* we cannot consider "our form of life as one among others." In other words, Kripkean skepticism is compatible with Lear's objectivism, in spite of the tension suggested by the terminology. The reason is that Kripke is skeptical about the possibility of meaning being given by an *interpretation* of a rule, and Lear's view does not depend on such a possibility. Indeed, what is unsatisfactory about Kripke's view is precisely that it ignores Wittgenstein's injunction to seek a "way of grasping a rule which is not an interpretation" (PI §201).

There is a second way of expressing the problem to which Kripke has drawn attention. It arises when we try to explain what it is that substantive disagreement over the application of a concept consists in. We are inclined to affirm a sentiment which we might sum up in the slogan, "No substantive disagreement without conceptual agreement," and which Bernard Williams has expressed more fastidiously as the need for an element in conflicting claims "which can be identified as the *locus* of exclusivity".[9] An apparent disagreement over a substantive issue will vanish if the parties concerned are arguing after all over the application of different concepts. But a problem arises here because of the part that agreement plays in the argument if Kripke's skepticism is allowed to stand, at least in the form he puts it. Consider the sentence that prepares the way for the famous "private language argument":

> If language is to be a means of communication there must be agreement not only in definitions but also (queer as this may sound) in judgments (PI §242).

The problem is this: if Kripke is right, then not only is agreement in judgments required for conceptual agreement, but agreement in judgments is *all* that conceptual agreement can consist in (he has put his skeptical scythe through everything else).[10] In that case *every* disagreement will be conceptual, and we shall never be able to say of two parties that they are in disagreement over some substantive issue rather than merely talking at cross purposes.

[9] "The Truth in Relativism," in *Moral Luck*, p. 135.

[10] A sign that this cannot be quite what Wittgenstein intended is given by his referring in the same paragraph to "what we call 'measuring' " as being only "*partly* determined by a certain constancy in results of measurement" (PI §242, my italics).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

This last point might be challenged. Surely, someone might say, Wittgenstein requires agreement in judgments only for primitive "ground-level" terms, the kind that are taught ostensively. But many terms are more complex, either those which are introduced and learned via definitions or theoretical terms about which we may legitimately disagree because we apply our principles of theory construction differently.

This reply will not do. Consider theoretical terms first (the argument is identical in the case of definitionally introduced terms). For example, take a disagreement over the truth of a scientific theory. If this is not due to a conceptual disagreement over observational terms, it must be due to a difference in application of the principles of theory construction. But what entitles us to talk about a different application of the same principles of theory construction? For Wittgenstein, as interpreted by Kripke, these principles just are the way in which we in fact construct theories from evidence. There can be nothing else for the meaning of these principles to consist in. But then it follows that two people who build different theories on the same evidence must be using different principles and, therefore, must be engaged in incommensurable tasks.

Consider a simple example. Suppose it were true that only primary-color terms (red, yellow, blue) were learned ostensively, and that secondary-color terms such as orange were always introduced by a definition such as "the color obtained by taking equal quantities of red and yellow (paint, say)." We may call the primary-color terms *first-order terms* and the secondary-color terms *second-order terms*. Now consider an altercation with someone who disagrees with us about which objects are orange in the following way: normally he agrees with us, but on Sundays and alternate Tuesdays he persistently calls "orange" objects that to us are so nearly red as to be almost indistinguishable from it, and he denies that colors any closer to yellow can be called orange at all. Now it is tempting to think that we can call this disagreement conceptual if he disagrees with us on Sundays and alternate Tuesdays about the application of the *primary* concepts "red" and "yellow"—something we can easily test. But suppose he doesn't. Then, we might say, since he has learned the application of the concept of equal quantities elsewhere than by color mixing, we must be in conceptual agreement with him, and he accordingly substantively wrong. Indeed, only yesterday we saw him mix a fine Black Velvet with impeccably equal quantities of Guinness and champagne.

"Not so fast!" he cries. "It is *I* who am really going on the same way as before. Your concept is one of schmequal quantities, which are

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

just like equal quantities most of the time, except that, unaccountably, on Sundays and alternate Tuesdays you persist in behaving quite differently. Since yesterday was Saturday, that explains why you didn't object to my Black Velvet."

The point need not be labored. If Kripke's skepticism is so destructive in a two-level argument, an induction makes it clear that it will be destructive in an argument of any complexity whatever. It should not be suggested that this sort of problem is unique to Kripke's interpretation. For clearly there may exist quite general difficulties in distinguishing cases where we try and fail to perform some action, to follow some rule, from cases where we are not really trying to perform *that* action at all.[11] Sufficiently general and predictable failure may no longer be interpreted as trying. Nevertheless, Kripke's arguments create problems of a particularly serious kind. Before we conclude, however, that our notion of meaning (particularly of what McDowell calls its "contractual" nature; see p. 325) lies in ruins, I want to point out one other consequence which shows that Kripke's account really cannot be right.[12] It is this: if we take the skeptical argument seriously, then it is quite obscure what use there is in appealing to intersubjective agreement. The destructiveness of the skepticism is unaffected by whether or not there is intersubjective agreement. Of course, this problem can appear to arise for Wittgenstein himself, at least for remarks of his taken out of context. Thus when he says "Hence it is not possible to obey a rule 'privately', otherwise thinking one was obeying a rule would be the same thing as obeying it" (PI §202), it is perfectly reasonable to ask, "How does intersubjective agreement help? Why is it less objectionable to say that, for the community as a whole, thinking it is obeying a rule is the same thing as obeying it?"

The question is entirely fair, but there is an answer to it which is quite consistent with what Wittgenstein says elsewhere. Kripke, however, cannot answer it; indeed the problem can be posed in his terms in a particularly stark form. For if there is no fact of the matter in virtue of which a particular action constitutes "going on in the same way as I did before," then likewise there is nothing in virtue of which

[11] Consider the oddity of the following case: a person announces that she is going to give a display of ventriloquism, and proceeds to talk with her lips moving energetically. A second person objects: "Surely ventriloquists aren't supposed to move their lips!" "Ah", replies the first, "I'm trying to be a *bad* ventriloquist."

[12] It is also an objection to Crispin Wright's account in his *Wittgenstein on the Foundation of Mathematics* (London: Duckworth, 1980), esp. ch. II.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

it is an instance of "going on in the same way as other members of my society." The fact that one comparison is simultaneous and the other diachronic is neither here nor there. For even if someone calls orange "red," this can be interpreted as an application of the same concept as that of his fellows if it is identical with their concept of red except in certain spatiotemporal locations, which only he occupies.[13] Faced with this, Kripke cannot even take refuge in an extreme sociological view, according to which society sometimes objects to my application of a concept and sometimes doesn't, and that is all there is to it. For how can we be sure that the utterance of a certain noise constitutes an objection? Surely, someone might say, it could be a "schmobjection", where . . .

The bizarre nature of this conclusion should help to alert us to what has gone wrong. What we want to say is that there *is* some fact of the matter with respect to intersubjective agreement, and Kripke is right (though inconsistent) in arguing as though there were. Of course, it is a "fact" relative to our perceptions of similarity and to our scheme of things, but since we have no notion of a fact that isn't likewise relative, this should not trouble us.

If this is so, we have no reason to flinch from saying that it is a fact that acting in one way rather than another is what constitutes going on the same way as before. There need be no metaphysically loaded implications of expressing the point thus. In particular, this conclusion is neutral regarding Kripke's repudiation of the truth-conditional view of meaning. All that is implied is that any conditions which establish that members of a community are in agreement with one another in their application of a term, can in principle be adequate to establish the conformity of an individual's behavior to his previous practice regardless of what the community does. This conformity is a fact in virtue of the structure of the customs and practices of that community, its "forms of life." This structure is one we can examine and, in principle, explain. It has causes and effects. Thus what constitutes applying the concept "orange" the same way as before just is to apply it to those things which have the causal

110147

[13] I think this point is fundamentally the same as one made by McDowell, though it is put somewhat differently. He writes (*op. cit.*, p. 34) that "if regularities in the verbal behaviour of an isolated individual, described in norm-free terms, do not add up to meaning, it is quite obscure how it could somehow make all the difference if there are several individuals with matching regularities." See also Simon Blackburn in "Rule-following and Moral Realism" (esp. p. 183), in Steven Holtzman and Christopher Leich, eds., *Wittgenstein: To Follow a Rule* (London: Routledge & Kegan Paul, 1981).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

structure characteristic of the things we have called "orange" in the past, that is to say, orange things.

Then it is in the investigation of the "structure" of our forms of life that social science consists. Far from being put out to grass by Wittgenstein, social science assumes a crucial role. And it is reasonable to understand Wittgenstein's continued emphasis on the fact that forms of life are what remain when explanations are exhausted, not as a spuriously formulaic "solution" to a difficult problem, but as a simple insistence that the answer is not to be found in an appeal to the mentalistic entities against which so much of the fire of the *Philosophical Investigations* is directed. Meanings are not in the head but in the world; but the emphasis is on the negative rather than the positive aspect of this claim. For to say that meanings are in the world is only the beginning of an answer. Nobody need imagine that it is fruitless (even for a strictly semantic inquiry) to try to understand the world a bit better than we do.

This might seem like an excessively glib dismissal of the skeptic. Kripke, after all, attacks the view that a rule "can be *embodied* in a machine that computes the relevant function" (p. 32 ff). It should be made clear that the important insight that motivates Kripke's attacks is preserved in the account given here. For the facts that constitute our forms of life, and thus our meaning one thing rather than another by the words we use, are not facts we can appeal to as a justification of *our* going on in one way rather than another—but this is because to do so is useless, not false. To say "I call this object 'orange' because it is orange" does not in the least strengthen my case for calling it "orange," even though there is another sense in which we want to say that the object's being orange *is* what justifies my calling it so. The way to sort out the ambiguity is to hold tightly onto the distinction between use and mention. Just as the sentence

(A) "Snow is white" is true iff snow is white.

says something important, whereas

(A′) "Snow is white" is true iff "snow is white" is true.

says something trivial; so the sentence

(B) To assert "This is orange" is justified iff my assertion is related in the appropriate way to the fact that this is orange.[14]

[14] I have not bothered to relativize the sentence, but to do so would change nothing. Nor have I filled out the "appropriate way" for an assertion to be related to what it asserts. This is not the point at issue.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

says something important, whereas

(B′) To assert "This is orange" is justified iff to assert "My assertion 'this is orange' is related in the appropriate way to the fact that this is orange" is justified.

is trivial.

Kripke's (and Wittgenstein's) insight is that to appeal to an interpretation of a rule (including an interpretation of a rule as embodied in a machine) in justification of *our own* way of interpreting a rule is entirely unhelpful, akin to the triviality of (A′) and (B′). Justifying an interpretation by appeal to another interpretation is just like thinking that we can give a general account of meaning in language by means of translation relations between languages—a theory of meaning of this form:

(C) "Snow is white" is true if "Schnee ist weiss" is true.

This, I think, helps to explain Wittgenstein's important remark that: "Hence there is an inclination to say: every action according to the rule is an interpretation. But we ought to restrict the term 'interpretation' to the substitution of one expression of the rule for another" (PI §201).

Unlike Kripke, however, Wittgenstein does not stop there. In the same paragraph he writes

> What this shows is that there is a way of grasping a rule which is not an interpretation, but which is exhibited in what we call 'obeying the rule' and 'going against it' in actual cases (PI §201).

Or, as we might put it, there is a way of giving the meaning of mentioned expressions which is not merely the substitution for them of other mentioned expressions, but of expressions in *use*.[15]

Seen in this light, the role of intersubjectivity in the argument becomes much clearer. This role is twofold. First, and most importantly, it arises from the split perspective that is required in the enterprise of explaining mentioned expressions via use. There is a sense in which for an individual to explain his own mentioned expressions via his own use is pointless; the perspective of use is different from that of mention, and the supposition of a private language (or, more generally, a private criterion of correctness in following a rule) is an illegitimate attempt to conflate the two.

[15] Cf. the introduction to Gareth Evans and McDowell, eds., *Truth and Meaning* (New York: Oxford, 1976).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

A helpful analogy may be the following:[16] a private language is impossible in just the sense in which private knowledge is. The reason for talking about knowledge is to emphasize that it must be more than just firmly held belief. It must be true, justified, caused in the right way, and so on. It is, therefore, absurd to suppose that there could be private knowledge, meaning beliefs that were such that only the individual holding them could realize that they constituted knowledge. For any belief he might have about the truth, justification, etc., of a particular belief would be just another belief, tantamount to a reaffirmation of the first one. And we might conclude, adapting Wittgenstein: "Hence it is not possible to have knowledge 'privately': otherwise thinking one knew something would be the same thing as knowing it".[17]

But of course we cannot go on from this to the skeptical conclusion that there is *no such thing* as knowledge, even though for us to be able to assert that a certain belief constitutes knowledge requires our holding certain beliefs ourselves. The point is that the perspective from which we talk of truth, justification, appropriate causal chains, etc., is different from the perspective of the belief in question. In just the same way, one may suggest, to draw from Wittgenstein a skeptical conclusion about meaning is absurd.

The upshot of all this is that the appeal to causal analysis and explanation of our practices and forms of life, far from being illegitimate as the opening paradox of this paper suggested, is the very stuff of a strong and useful social science. In a real way it can display the facts in which our meanings consist. And we can easily explain Wittgenstein's apparent hostility to natural history as the resistance to a bootstraps attempt to justify one perspective in terms of itself. This becomes very clear in the discussion between paragraphs 169 and 180 of the *Philosophical Investigations,* where he mocks our tendency to "imagine that a feeling enables us to perceive as it were a connecting mechanism between the look of the word and the sound

[16] I am not yet confident that this analogy is entirely apt. But this is chiefly because the analogous version of Kripke's argument is so obviously specious, that I am reluctant to think that Kripke would attribute to Wittgenstein the ability to be taken in by its close relative.

[17] This analogy may help to make clear why interpretations of the private-language argument in terms of the impossibility of Robinson Crusoe's developing a language miss the point. Nothing in the private-knowledge argument would prevent us from saying that Crusoe had knowledge of his island.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that we utter" (§170). What he is mocking, of course, is not the belief that there are causal connections, but our appeal to them to justify the things we "just do."

There is a second, weaker part played by intersubjectivity in the Wittgensteinian "solution" to these problems. On the account I have given, there are indeed conceivable circumstances in which we might wish to say that the entire community of language users had "gone wrong," had betrayed, as it were, the understanding implicit in its earlier forms of life. We might make this judgment if these forms of life were of such a kind and had such a structure that to go on in the same way would have involved doing something different from what the community actually did—but, we might suppose, peculiar and extraneous causal factors intervened to prevent this from happening. An extreme example would be the following. Suppose that the accidental release of a certain chemical into the atmosphere were discovered some months later to have affected the retinas of the world's population so that certain pairs of colors, previously distinguishable, had become indistinguishable and vice versa. Then, we may imagine, everyday linguistic usage would have changed (and verifiably so) in that different color terms would systematically have been applied to given wavelengths of light. We might wish to say that the meaning of these terms had changed. But, more probably, we would describe this discovery as showing that the community had erred (indeed, when this was pointed out, the community might well correct its practice). Such an account makes perfect sense. Much political argument can be viewed in precisely this light (dissidents often argue that the community of which they are members has, either directly or through its political representatives, betrayed its prior intentions and commitments).

Nevertheless, as a pragmatic matter we surely tend to be rather cautious about accepting stories of this kind. The fact that the entire linguistic community has behaved differently from what we expected is likely to alert us to the possibility that we have misunderstood its original practices and to prompt us to re-examine them more closely. And likewise, when one individual's practice diverges from that of the community at large, explanatory parsimony inclines us toward suspecting him of misapplying his concepts. As Kafka put it, "In your struggle against the world, back the world."

It may even be, in such cases, that the structure of the individual's practice before the divergence proves to have been sufficiently different from that of his fellows (and in a way that can account for the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

divergence itself) for us to want to say that he was in conceptual rather than substantive disagreement with them.[18]

But, as is by now well established in the philosophical discussion of interpretive constraints on meaning by Donald Davidson and others, we may expect to find a general presumption in favor of diagnosing disagreements as substantive rather than conceptual. This is because, in order to establish the presence of a conceptual disagreement at all it will be necessary to make assumptions about other shared patterns of behavior. Thus suppose, for example, that in a certain society one member believes that goats yield wool. His fellows may believe him to be simply mistaken, or they may suspect that he is confusing goats with sheep, and thus is confused over the application of the concept "goat." But in order to diagnose the disagreement in the latter fashion they would have to assume that he agrees with them about the application of other concepts, such as "wool" (and "yield"). Nevertheless, this does suggest that it would be unwise to take our earlier slogan ("no substantive disagreement without conceptual agreement") too seriously. Just where we choose to say there is substantive and where conceptual disagreement is something governed by rather unclear principles, and will tend to reflect our optimism or pessimism about the prospects for resolving such disagreement.

Where does the discussion so far leave us in relation to the alleged paradox with which this paper began? In a simple sense, it has been claimed that assumption 1 does not entail assumption 1a and that, therefore, the conclusion 3 does not follow from assumptions 1 and 2. Contrary to (1a), explanations of our forms of life are not only possible but necessary. Sometimes these explanations will be directly causal in nature; at other times we shall wish to say that certain normative regularities of behavior occur because they are prescribed by *other* normative practices, so that they are directly comprehensible even before "explanations . . . come to an end."

[18] It is worth observing, parenthetically, that if we believed that the terms in which we explained and described forms of life were the only terms for describing the motions of human beings—if, that is, we subscribed to a rather crude kind of psychological determinism—then it would follow trivially that, if behavior subsequently diverged, its origins could not have been the same, and thus that all disagreement was by definition conceptual. The fact that we do not think like this, that we can accept the view that the influence of pactices and customs does not by itself *determine* future applications but can be overridden by causal factors at a different level, makes this account fit in quite well with an antireductionist attitude toward the mental.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

But this does not mean that there is nothing to be learned from the skeptical account given by Kripke. On the contrary, the fact that "explanations . . . come to an end" and that the explanations that occur, as we might put it, "afterwards" are from a different perspective (that of *use* rather than *mention,* in the terms of the discussion above), can help to illuminate some of the most vexed questions in the whole of social science. One example concerns the precise nature of the relation between the "internal" and "external" perspectives on action. More bluntly put, when do causal explanations of human customs and practices debunk what they explain? On the skeptical account in its simple form, the answer to this is a mystery, and even Lear's objectivism seems to put causal explanations entirely beyond the horizon. On the account outlined here, however, the fact that the causal structure of forms of life can, from one perspective, *constitute* our understanding, may help to illuminate this difficult relationship. For instance, what light might sociobiological theories shed on our traditional moral rejection of incest? The view I am suggesting requires us, if we are consistent, to affirm that our rejection of incest is relevantly similar to our established practices of rejecting torture, murder, and other less controversially immoral acts. A sociobiological account can indicate that, in some respects, our practice of rejecting incest may have an importantly different structure from these others. It is essential, however, not to misunderstand this point. We can only say "*may* have." For, as the skeptic in Wittgenstein and in Kripke points out, if someone insists that, nevertheless, there *is* a relevant similarity of structure between our incest taboo and other moral practices, there is nothing we can do to prove him wrong. But it may be, however, that, once we can show a plausible dissimilarity, his faith in their similarity may either falter or be required to locate the similarity in ever more ineffable properties of these practices. Thus the link between external and internal perspectives can never be direct or compulsive, but is not negligible for that. It is a link that is, in *one* sense of a misleading word, psychological. Causal explanations can never *require* us to alter our practices (as the Kripke view of Wittgenstein importantly reminds us)—but they can help to illuminate them, and even to weaken the hold that some of our more unreflective forms of life have upon us.

PAUL SEABRIGHT

All Souls College, Oxford

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

A PHILOSOPHICAL LETTER OF ALFRED TARSKI

Prefatory Note. This letter from the late Alfred Tarski to me in September of 1944 should be of considerable historical interest to many philosophers. Unfortunately, I cannot find a copy of the letter I wrote to him which prompted his reply. I do recall, however, that I had asked him some questions about his reaction to certain passages in Felix Kaufmann's *Methodology of the Social Sciences.*[1] Kaufmann had implied on page 66 of that work that "the meaning of ['true'], if applied to analytic propositions, is essentially different from its meaning when applied to synthetic propositions," and Tarski addresses himself to this sort of view in the earlier parts of this letter.

Later in this letter, however, Tarski discusses the circumstances under which he was ready to reject the logical and physical premises of a science. Here his views are similar to views later advanced by W. V. Quine in his "Two Dogmas of Empiricism".[2] In the first footnote of that paper, Quine acknowledges a "large and indeterminate debt" to Tarski and others, and Tarski writes in this letter: "I think that I am ready to reject certain logical premisses (axioms) of our science in exactly the same circumstances in which I am ready to reject empirical premisses (e.g., physical hypotheses); and I do not think that I am an exception in this respect." By 1940, Quine reports in his autobiography,[3] he had developed doubts about the notion of an analytic sentence. He also reports there (p. 150) that Tarski had joined him in these doubts when both of them were members of a discussion group to which Carnap tried, unsuccessfully, to read the whole of the manuscript of his *Introduction to Semantics.*

I should add that I had letters such as this one from Tarski in mind, as well as conversations with him, when I wrote, in the opening footnote to my article "The Analytic and the Synthetic: An Untenable Dualism,"[4] that my general attitude on the issue had been influenced not only by Goodman and Quine but also by Tarski.

For the reader's further information, I add that the "article on logical consequence" to which Tarski refers in his letter was later translated into English as "On the Concept of Logical Consequence",[5] and that "the recent one on truth" to which he refers is "The Semantic Conception of Truth and the Foundations of Semantics."[6]

I am grateful to Tarski's widow, Maria J. Tarski, for agreeing to the publication of this letter.

MORTON WHITE

The Institute for Advanced Study

0022-362X/87/8401/0028$00.50
© 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

UNIVERSITY OF CALIFORNIA

Department of Mathematics
Berkeley, California

2606 Shasta Rd.
Berkeley 8, Calif.
September 23, 1944

Dear Morty:

It is very difficult to discuss in writing the problems you have raised in your last letter. And, as a matter of fact, I cannot go much beyond what I have said in my article on logical consequence and in the recent one on truth (notice, in particular, my remarks in §14). But I shall try to be more specific.

Let us not forget even for a minute that any definition of 'logical term' and 'logical truth' can only be given in terms of a determined language and in reference to a determined language (or to a determined class of languages). It is clear that for all languages which are familiar to us such definitions can be given (or rather: have been given); moreover, they prove fruitful, and this is really the most important. We can define 'logical terms', e.g., by enumeration; but you realize that there are always infinitely many equivalent forms in which the same definition can be put. It is true, we can consider even the possibility of several non-equivalent definitions of 'logical terms' ('log.$t.^{(1)}$', 'log.$t.^{(2)}$', etc.); e.g., sometimes it seems to me convenient to include mathematical terms, like the ϵ-relation, in the class of logical ones, and sometimes I prefer to restrict myself to terms of 'elementary logic'. Is any problem involved here?

---

[1] New York: Oxford, 1944.

[2] *Philosophical Review*, LX, 1 (January 1951): 20–43; see especially pp. 39/40. Reprinted, with changes, in Quine's *From a Logical Point of View* (Cambridge, Mass., Harvard, 1953); see especially pp. 42/3. This essay has been reprinted elsewhere.

[3] *The Time of My Life* (Cambridge, Mass.: MIT Press, 1985).

[4] *John Dewey: Philosopher of Science and Freedom*, Sidney Hook, ed. (New York: Dial Press, 1950), p. 316, n. 1. Reprinted, with changes, in my *Pragmatism and the American Mind* (New York: Oxford, 1973), especially pp. 121/2, n. 1, and elsewhere. In this note I mention the importance of taking into account John Stuart Mill's role in "the history of discontent with the distinction between the analytic and the synthetic". Tarski, as the reader will see, has occasion to mention the impact of Mill on his own thinking about this and allied topics.

[5] The English translation by J. H. Woodger appears in Alfred Tarski, *Logic, Semantics, Metamathematics* (New York: Oxford, 1956), pp. 409–420. Also see the *Bibliographical Note* attached to this article on p. 409.

[6] *Philosophy and Phenomenological Research*, IV, 3 (March, 1944), especially pp. 355/6.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

It is also clear that we can define the notion of a logically true sentence for all familiar languages. We can do this in many different ways. Let *Ls* be the class of logical sentences, *Es* that of extra-logical sentences, *Lt* the class of true logical sentences, *Et* that of true extra-logical sentences, and *Tr* that of all true sentences. Then we can find it convenient either to define first *Tr* and then *Lt* and *Et* as intersections $Ls \cap Tr$, $Es \cap Tr$, or else first *Lt* and *Et*, and then *Tr* as the union $Lt \cup Et$. Both procedures are *always* possible, independent of the contents of definitions. I can only add that in the semantic conception of truth (and this is, in my opinion, the only precise conception of truth which is now available) it seems to me 'simpler' and 'more natural' to define first '*Tr*', and then, with its help, '*Lt*' and '*Et*'; but I don't use here the words 'simpler' and 'more natural' in any precise sense.

By the way, if by a logical sentence you understand a sentence which contains no extra-logical terms, then the definition $Lt = Tr \cap Ls$ would not seem convenient and fruitful to me. I'd prefer a definition which would comprehend also all logical consequences of such sentences, i.e., results of substitution, of replacement of definiendum by definiens, etc.

To the problem whether 'true' as applied to *Ls means* something different than 'true' as applied to *Es:* the answer depends on how you define the expression '*A* means in *B* something different than in *C*'. I must say that I simply don't know what such a phrase means. I imagine a useful definition of meaning in contexts like '*B* and *C* (i.e., two sentences or expressions) have the same meaning', 'the meaning of *B*' etc. But 'the meaning of *A* in *B* is different from that of *A* in *C*'—this seems to me too involved. Of course, I don't mind if anybody defines this phrase in whatever way he wants and gets whatever answer he wishes.

Are the sentences like 'Whatever is colored is extended' *necessary?* This depends again on your definition of necessary sentences and of the language to which you apply.

You may think that I am joking the whole time. But this is not the case. I really don't understand what problems are involved here; and if there are any, they certainly must be formulated in an entirely different way. I have noticed the statement in Kaufmann's book to which you refer; but I seriously think that Kaufmann (like most other 'philosophers') does not have enough 'logical culture' and does not think clearly enough to discuss problems of this type.

I repeat that if any problems or points of difference are involved here, they must be formulated in a different way. At any rate, I would not feel able to participate in any discussion concerning the 'legality'

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of definitions, 'essence' of the notions defined, etc. The situation would change if problems were formulated in a sound, reasonable way, in terms of logic, physics, psychology, etc. Here are some samples of such formulations (of course, very rough samples).

I would be inclined to believe (following J. S. Mill) that logical and mathematical truths don't differ in their origin from empirical truths—both are results of accumulated experience. A rough example. In a very early stage of their development, people learned to use the words 'not' and 'or'. In certain cases they were sure that something was white, in other cases that it was not white. In many cases they were first unable to decide whether a given thing was white or not (e.g., as a result of a bad light). But they noticed that in many such uncertain cases they finally reached a decision—by means of more thorough and repeated observations, better instruments, etc. Hence they began to believe in 'Everything is white or is not white' and, more generally, in '*p* or not *p*'. Of course, the whole problem belongs to the history of science, of human thought, and is not of a 'philosophical' nature; it may happen that I am completely wrong. In addition, the problem has no fundamental importance.

Another, more important problem, of a psychological character. I think that I am ready to reject certain logical premisses (axioms) of our science in exactly the same circumstances in which I am ready to reject empirical premisses (e.g., physical hypotheses); and I do not think that I am an exception in this respect. Explanation: We reject certain hypotheses or scientific theories if we notice either their inner inconsistency, or their disagreement with experience, or rather with individual statements obtained as results of certain experiences. No such experience can logically compel us to reject the theory: too many additional hypotheses (regarding the 'initial conditions', circumstances of the experiment, instruments used) are always involved. We can practically always save the theory by means of additional hypotheses. But often these additional hypotheses are so involved and 'unnatural' that they make us uncomfortable—we begin to feel an inner need of rejecting the theory or some parts of it, and of replacing it by a new, simpler and more natural theory, which would not require these additional hypotheses. It depends on the nature of experience what we reject—a rather special law which is an 'inductive generalization' of individual statements, or a more general and profound hypothesis, or even one of fundamental premisses of our science (e.g., the rejection of Newton's mechanics or Euclidean geometry). Axioms of logic are of so general a nature that they are rarely affected by such experiences in special domains. However, I don't see here any difference 'of principle'. I can imagine that cer-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tain new experiences of a very fundamental nature may make us inclined to change just some axioms of logic. And certain new developments in quantum mechanics seem clearly to indicate this possibility. That we are reluctant to do so is beyond any doubts; after all, 'logical truths' are not only more general, but also much older than physical theories or even geometrical axioms. And perhaps we single out these logically true sentences, combine them in a class, just to express our reluctance to reject them.

Whether this description is true and adequate—I don't know. I have the impression that many people would agree with it. I think, e.g., that Carnap would agree. Nevertheless, he would claim that there is a fundamental difference between a change of logic and a rejection of a physical theory; for, he would say, in the first case we change the language, we begin to use the words in a new meaning, while in the second no such change occurs. Why? Well, this follows from his definition of meaning (which in turn is based on a definition of logical terms and logical truth). Of course, he is permitted to do so; he can accept whatever definition of meaning he wants; and perhaps I would find it also convenient to accept his definition. However, I don't see how such a definition can really affect the situation; if you define logical axioms as those which cannot be changed without the change of your whole language, then of course you cannot change them without changing the language—this is a truism (at least, in our present logic).

So much about the problems in which you are interested; I doubt very much whether you will find these remarks satisfactory.

I must apologize for not answering your earlier cards. But you can easily imagine how badly I feel during these last months—in connection with the situation in Warsaw. I have of course no news from my family since the uprising and the siege began, and I am afraid that even if they survive this ordeal I shall not hear from them for a long time. Russians are not anxious to permit their subjects to communicate with the rest of the world. I have not yet received any word from those parts of Poland which were "liberated" many months ago, though I have some very good friends there who know my address.

In this situation I can hardly do anything but worry and listen by radio; even to write a letter is for me an effort. I hope I'll be a better correspondent if the situation changes.

With best regards to you and Lucia.

Cordially
Alfred

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

Mathematics: Form and Function. SAUNDERS MACLANE. New York: Springer-Verlag, 1986. 476 p. $42.00.

This book, by an eminent mathematician and coinventor of category theory, urges richer interaction between the philosophy and the practice of mathematics than we have seen in fifty years. It illustrates as much as states a philosophy of mathematics, and it lays out the background MacLane believes the philosophy of mathematics requires. He makes high demands, and has a high opinion of what philosophy of mathematics could be if it drew on the actual wealth of mathematics. He finds too many theories of foundations "are anchored almost exclusively in the most elementary parts of Mathematics—numbers and continuity" (4), and deprecates Wittgenstein's work "where the actual mathematical content rarely rises above third grade arithmetic, while the actual concern is less with Mathematics than with its use to illustrate some purely philosophical issue" (444).

So most of the book presents various branches of mathematics. These clear, extensive descriptions include well motivated, nearly rigorous proofs for an array of major theorems from linear algebra to complex analysis. The book is not entirely self-contained. MacLane assumes the reader "has at hand some of his own familiar texts for possible reference" (5). But a reader who knows a bit of set theory, groups, linear algebra, and calculus will learn a great deal; and one who knows more will learn more.

The book may attract philosophers by offering the most accessible introduction to yet category theory and topoi. This section has no special prerequisites, although the examples draw on other subjects. MacLane gives detailed definitions of the basic tools of category theory and of all the terms used in the topos axioms. He gives examples of topoi based on sets. He does not advocate the radically categorial foundations of mathematics associated with F. William Lawvere, who has worked with him.

MacLane's basic example of a topos is the class of all sets posited by ZF set theory with Choice, plus the functions between them. But, as he notes, a general topos need not be much like a universe of classical sets. So he immediately adds axioms for a well-pointed topos with natural numbers and Choice. These axioms are essentially a categorial version of Zermelo set theory, and MacLane offers them as alternative foundations for mathematics. He says this approach "has nowhere been carried out in great detail" (402), but this is so

0022-362X/87/8401/0033$00.50 © 1987 The Journal of Philosophy, Inc.
CC-0.In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

close to set-theoretic foundations that there is a fairly routine translation.[1] He ignores other extensions of the topos axioms which are inconsistent with classical set theory but give neatly tailored foundations for specific branches of mathematics, such as axioms for synthetic differential geometry.[2]

MacLane is also less committed than Lawvere to categorial methods in the practice of mathematics. He takes the fact that "categories and functors are everywhere in topology and parts of algebra, but they do not yet relate well to most of analysis" (407) as showing a weakness in category theory. Lawvere takes the same fact as showing the infelicity of those parts of analysis most affected by set-theoretic technicalities. Lawvere does not take categorial foundations as the last word either, but as a more powerful latest word than does MacLane.

MacLane is nevertheless a major architect of the categorial attitude. It shows throughout his survey of mathematics, even in the wide range of material he takes as essential to philosophy of mathematics. If mathematics were only arithmetic and basic calculus there would be no use for categories or categorial foundations. Samuel Eilenberg and MacLane invented category theory in the early 1940s as a tool for homology theory. Many central theorems of topology are proved by assigning to each topological space various groups, called its *homology groups*. In 1940 this powerful method was accessible only to insiders, and its basis was obscure even to them. Category theory clarified the situation, so homology theory became accessible to any serious mathematician and was itself fruitfully extended. Category theory arose from a practical need to relate two great subjects: group theory and topology, a need remote from traditional interests of philosophers of mathematics.

Chapter Five typifies MacLane's style and purpose in surveying mathematics. It begins with conventional set-theoretic treatment of functions plus a mention of category theory, then passes through transformation groups to a clear five-page sketch of Galois theory. It describes some constructions on groups and then the finite simple groups. The final section asks why group theory is so rich in comparison to other axiomatic theories. This section draws on all the themes I have listed and more; yet it can be read as quite banal unless the reader actively applies it to the earlier material. MacLane only opens the question and gives a framework for answers.

[1] This should be clear from the discussion of the category of sets in W. S. Hatcher, *The Logical Foundations of Mathematics* (New York: Pergamon Press, 1982).

[2] Synthetic differential geometry was conceived by Lawvere and can be seen in A. Kock, *Synthetic Differential Geometry* (New York: Cambridge, 1981).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

On calculus MacLane gives scant attention to the well-worn topic of infinitesimals versus limits, and stresses topological ideas and relations to physics. The chain rule quickly leads to gradients, tangent spaces, cotangent spaces, and duality. This may be more than the reader wanted to know about the chain rule, but later it is crucial to MacLane's own geometric derivation of Hamiltonian dynamics. Then he discusses tricks, such as the apparently ad hoc substitution of variables usually used to derive Hamiltonian dynamics, versus ideas, such as the geometric ideas he claims are actually behind the derivation. This gives a more accurate idea of practice in calculus and differential geometry, and a more accurate sense of their history, than the common fascination with Dedekind cuts and ϵ-δ definitions.

The introduction and the last chapter are more purely philosophical, with some twists. MacLane asks about the wealth of mathematics surveyed: "How does it illuminate the philosophical questions as to Mathematical truth and beauty and does it help to make judgements about the direction of Mathematical research?" (409) He knows that few philosophers ever ask about mathematical beauty and few mathematicians care for judgments about the direction of research informed by a philosophy of mathematics. But he studied mathematics at Göttingen 1931–1933. He heard Emmy Noether lecture on algebra, and Hermann Weyl on geometry, algebra, and the philosophy of mathematics. He stayed in Weyl's house for a time and wrote up Weyl's lectures on philosophy of mathematics. He did a dissertation on logic under Paul Bernays. All this in an atmosphere shaped more by the recently retired David Hilbert than by anyone else. In Göttingen philosophic debate over formalism, logicism, and intuitionism had been a part of mathematical research for some decades. And if "On the Infinite" is usually cited as a source of ideas on metamathematics, remember that Hilbert wrote it as a defense of beautiful mathematics—of "paradise." MacLane knows how far mathematics has come since then and has no desire to revive old philosophies; but he keeps the standards he learned in those days.

He summarizes his "formal functionalism" this way: "Mathematics aims to understand, to manipulate, to develop, and to apply those aspects of the universe which are formal" (456). By "universe" MacLane means all there actually is around us. So this is not a philosophy of pure mathematics leaving applications as another question. And yet he says that mathematics is irrefutable. It is irrefutable because it deals with forms and not facts; yet it is applicable because it deals with forms of facts (414). On the other hand, he says that any truth must be falsifiable in experience, so "Mathematics is not true, but its correct results are certain" (442). This is not the Kantian doctrine

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that we know truths about form a priori. Nor is it positivist, since MacLane treats forms as being objectively in the universe and formal knowledge as far from meaningless. Throughout the book he points out alternative proofs for various results, then singles out one as giving the real reason for each theorem (145, 189, 427, 455).

MacLane says a number of things about form. He draws on earlier chapters to show how many ways the basic ideas of geometry have been formalized and how each formalization has unforeseen applications. He says no single formalism such as set theory or category theory can serve as foundation for all mathematics, because it would "preclude the novelty which might result from the discovery of a new form" (455). And yet mathematics cannot proceed without formalizing its ideas: "Before that an idea can be pretty nebulous; it is 'up in the air.' That may be why it is hard to convey to others some new intuition about Mathematics" (444). The remarks, though only suggestive, are richly suggestive in the context of MacLane's contribution to mathematics and formal foundations.

He claims, since mathematics can be correct but cannot be true, that it makes no ontological commitments and there is no epistemological question of mathematical knowledge. There is only the question of how to recognize correct proofs, and this is internal to mathematics. "But this does not dispose of the hard questions about the philosophy of mathematics; they are merely displaced" (444). They are removed from epistemology and ontology and embedded firmly in the practice and history of mathematics. He raises such questions as "What are the characteristics of a Mathematical idea?" and "How does Mathematical form arise from human activities or scientific questions?" and argues they must be approached with specific, often advanced, knowledge of mathematics (444–446).

It may be jarring to find, twelve pages from the end of a lengthy book, that the hard questions have been not solved but only displaced. MacLane does not begin to answer all the questions he raises and does not finish the answers he begins. He aims to open up a field neglected since the Nazis ended the great days of Göttingen—philosophy of mathematics as a whole, mathematics at its best, not only everyday arithmetic and the basics of calculus, and not only advanced set theory or meta-mathematics. It is an exciting try. Nonetheless a philosopher reading this book has to bridge a cultural gap between philosophy and mathematics. Such a gap depends more on different experiences of learning and research than on specific differences in subject matter or technique. It must be evoked rather than described. This one is epitomized in a phrase which runs between the lines of MacLane's book without appearing in print, a phrase indis-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pensable in mathematics though it would read as a joke in most philosophical writing: "Details of the proof are left to the reader."

COLIN MCLARTY
Cleveland State University

Language and Logic in Ancient China. CHAD HANSEN. Ann Arbor: The University of Michigan Press. 1983. 207 p. $25.00.

This book has, in a sense, fallen through a crack. Analytical philosophers have not picked it up because, by and large, they are not interested in Chinese philosophy; and the more traditional intellectual historians and others who *are* interested in Chinese philosophy have not picked it up because they have been put off by the tight, aggressive argumentation Hansen marshals and the bold, even iconoclastic, interpretations he introduces and defends. Yet the book has something to offer each camp. For the modern philosopher, Hansen's sophisticated interpretations and careful comparisons with modern theories of meaning and reference bring a number of neglected Chinese classics back to life by making apparent their genuine philosophical interest. For the more traditional scholars of Chinese thought, Hansen's book presents a strong methodological challenge; for to defend the traditional interpretations against Hansen's alternatives, a new approach will be needed: Hansen's arguments will have to be met head on, and at a level as least as rigorous, analytical, and philosophically informed as Hansen's.

The book begins with a chapter called "Methodological Reflections" in which Hansen presents his ideas on how to develop and evaluate an interpretation (and translation) of a text. Hansen's views owe much to W. V. Quine's theory of radical translation, but he also emphasizes Richard Grandy's "principle of humanity" according to which people will make roughly the same sorts of inferences if they start from the same assumptions. Thus, for Hansen, it will not be enough for an interpretation to make a text look more or less true; we must interpret it in a way consistent with our understanding of an author's assumptions, motives, and background knowledge generally. On one level the book can be seen as an attempt to present such a theory of interpretation and illustrate how it works. In Chapter Five, for example, Hansen develops an interpretation and translation of Kung-sun Lung's "White Horse Paradox" in which Kung-sun Lung is taken as saying something which, although false, is neverthe-

0022-362X/87/8401/0037$00.50 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

less an intelligent claim if understood as a counterexample to certain Mohist claims that Hansen thinks Kung-sun Lung was trying to argue against. In contrast, the more traditional interpretations have Kung-sun Lung saying something which, although true, is trivial and in Hansen's view bears no relation to anything else in Chinese philosophy of language of the period. Of course, to support these claims, Hansen must take up the main theories of language and logic in the classical period, and this he does *inter alia* in Chapters Two, Three, and Four.

In Chapter Two, "The Mass Noun Hypothesis and Abstraction in Chinese Language and Thought", Hansen presents some of his most controversial ideas, although he begins with some fairly straightforward observations about classical Chinese grammar. Chinese nouns, as is well known, function for the most part without inflections such as singular and plural, much like the "mass nouns" in English such as 'water', 'furniture', and so on. We cannot say "one water" or "two furnitures," although we can say "one cup of water" and "one piece of furniture." It is as if the word 'furniture' referred to a single, spatially discontinuous object, and by saying "this piece of furniture" we pick out this one part of a larger object. In this way Hansen views mass nouns (and Chinese nouns generally) as functioning as singular terms or names: the individual picked out by a noun is going to be a certain *stuff* (water, furniture, horse, or whatever). Of course furniture, unlike water, is not a stuff in any straightforward intuitive way, but Hansen wants to suggest that anyone with a knowledge only of classical Chinese (i.e., anyone in classical China) would have been led by the grammar of the language to view the whole world as a collection of stuffs. The stuffs might overlap or interpenetrate, but the basic individuals in the world would have been thought to be those various stuffs.

The problem with this view, one could suggest, is that the difference between viewing a chair, for example, as a discrete object and viewing it as a part of a certain discontinuous sort of furniture stuff is not very clear. A strong Quinean, for example, might argue that, for a given speaker, there is no fact of the matter about which view is correct, and indeed no empirically verifiable difference between them. Yet, with no further ado, Hansen goes on to make an even more controversial claim, the claim that with an ontology of stuffs there would have been no reason for a classical Chinese philosopher to posit an *abstraction*—such as a Platonic Form, a Lockean abstract idea, or a Fregean sense. In particular, whereas according to traditional western semantics a word refers to an object because the latter instantiates the "abstract idea" associated with the word, Hansen

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

claims that no such abstractions would have been required or even thought of in classical China, Chinese nouns functioning as they do as mass nouns akin to names. (On this, more below.)

In Chapter Three, "Background Theories of Language in Ancient China", Hansen tries to discuss various points in Lao Tzu, Confucius, Mo Tzu, Chuang Tzu, and others, with a view to showing that none of these figures postulated abstractions or even presupposed them. His aim here is to defend his theory of the nominalist ontology he thinks existed in classical China and, thus, to strengthen his view that there was no precedent for Kung-sun Lung to be talking about the universals and abstractions that have figured so centrally in the traditional interpretations and translations of Kung-sun Lung which Hansen opposes.

In the fourth chapter, "Neo-Mohist Philosophy of Language", Hansen takes up, in somewhat scholastic detail, a number of fragments from the Neo-Mohist writings, particularly the Mohist Canons I and II, the Canon Explanations I and II, and two obscure Mohist fragments known as the "Greater Pick" and the "Lesser Pick." The aim here is to show that these fragments too are consistent with the characterization of classical Chinese language Hansen has developed so far and to show, furthermore, how some of these fragments might be interpreted as something Kung-sun Lung might have been interested in arguing *against*.

The various strands of the book come together in Chapter Five, "Kung-sun Lung and the White Horse Paradox", where Hansen presents his interpretation of the White Horse Paradox and defends it by reference to the various points he has developed thus far. Rather than interpret Kung-sun Lung's famous "*pai me fei ma*" (literally, "white horse not horse") in terms of universals and abstractions, Hansen does it (somewhat obscurely) in terms of stuffs, stuff-parts, and word mentions.

Now, to evaluate Hansen's interpretation, one approach, clearly, would be to look more carefully at his theory that classical Chinese presupposes a nominalist ontology; another would be to look at the relation he thinks he sees between Kung-sun Lung and the Neo-Mohists. If either side falters, then Hansen's position would be, if not refuted, at least undermined.

I shall not take up the question of Kung-sun Lung's relation to the Neo-Mohists, except to suggest that several of the texts concerned are so obscure, even corrupt, that a bit of skepticism might well be called for, no matter what the interpretation. Instead, I will take up some of Hansen's comments on nominalism and abstraction, for these tie in well with main themes in modern philosophy of language

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and will be of some interest even to readers who lack a specialized knowledge of classical Chinese.

According to Hansen:

> Naming [in classical Chinese] is not grounded on the notion of an abstract concept, a property, an essence, or an ideal type, but rather on finding "boundaries" between things. Accordingly, Chinese philosophers view minds not as repositories of weird objects called ideas, but as the faculty encompassing the abilities and inclinations to discriminate stuffs from each other (32). . . . This "cutting up things" view contrasts strongly with the traditional Platonic philosophical picture of objects which are understood as individuals or particulars which instantiate or "have" properties (universals) (30).

But this leaves us with the problem of understanding how the non-Platonic, "cutting up things" Chinese view is supposed to work in detail. Hansen attempts to explain this by telling a quaint story about how Baby Susie learned the word 'doggie' and Baby Mei-ling the word '*kou*'. For Baby Susie, it is the familiar Lockean model of abstract ideas: "she has abstracted from all the particular dogs she had encountered the features common to all dogs" (52). But for Baby Mei-ling the story is quite different: "the story . . . does not involve any abstracting. Rather one says that she has acquired the ability to distinguish dog-stuff from non-dog-stuff. She is, in effect, not seeing a different object, but a different part of the same stuff" (52). The problem with this answer is that, whereas the Lockean model, despite its well-known problems, provides an account of name learning that is at least superficially plausible, the Mei-ling model does not seem to be providing any coherent account at all: all that is said, really, is that Mei-ling learns to name by "discriminating" without abstracting—which evades the question of *how* Mei-ling is to identify or discriminate those different parts of the same stuff when she applies the word '*kou*'. Having the "stuff ontology" that Hansen describes can hardly make such identification unnecessary since, surely, and as Hansen realizes, Baby Mei-ling will want to distinguish one dog (or hunk of dog-stuff) from the next. How does Baby Mei-ling's stuff ontology make it any less likely that she would make this distinction by remembering properties true of certain parts of dog stuff but not others—and isn't this what abstracting comes to? If so, it would seem wrong for Hansen to insist that describing Mei-ling's problem as learning to reidentify the same stuff "makes us less likely to talk of abstracting properties from different objects" (52). Even if Hansen is right that classical Chinese philosophers did not speak of abstraction in their theories of language, the Baby Susie/Baby Mei-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ling case hardly seems to provide an explanation of why this might be so. Perhaps the idea just didn't cross the minds of classical Chinese philosophers; or, more likely, perhaps they simply didn't share the western obsession with explaining how a word picks out its referent.

Elsewhere Hansen gives another account of how he thinks the identification of individuals (or the identification of disparate "stuff-parts") might be supposed to work for a classical Chinese speaker if abstraction theories are out. Ironically, he thinks the allegedly pictographic Chinese written characters function for a speaker of classical Chinese the way abstractions function for a speaker of English:

> The Chinese can rely on picturing or representing as the method of tying language to things without having to invent a detour through the mind and mental images. The picture is itself the linguistic entity, the character. The written form is not a representation of the sound but of the thing. Hence the mediation between sounds and objects in the world is not provided by the inner, private subjectivity (an idea or other "affection of the soul") but by shared social convention—by the character (49).

But this theory cannot be right. In the first place, although Chinese is sometimes described as "ideographic," anyone familiar with the Chinese character for dog can attest that it no more resembles a dog than the English word 'dog' does. Even if the character somehow consisted of an actual photograph of a dog, this would not help a Chinese speaker apply the word '*kou*' unless he first knew that dogs needn't all have that exact shape, color, and so on. So at best much else would be required beyond the pictograph. In the second place, Hansen's theory would seem to entail that Baby Mei-ling can distinguish one dog from the next only *after* learning the written character, which is absurd. Since young children learning Chinese almost invariably come to use words like '*kou*' correctly well before acquiring any knowledge of the written language, the written language can hardly be the feature that best explains how this word use occurs.

Hansen may yet be correct in his wider claim that classical Chinese philosophy lacked theories of abstraction, but the implausibility of his nominalist theory of language learning will hardly help his case or deter his critics from the traditional view of the White Horse Paradox as somehow presupposing a theory of abstraction, such as a theory of universals.

On the other hand, Hansen has many other arguments up his sleeve. Many of these have intrinsic interest far transcending whatever relation they may have to the White Horse Paradox. Indeed, the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

importance of this book probably lies less in the contribution it makes to Kung-sun Lung scholarship (Kung-sun Lung being, after all, a comparatively minor figure in the history of Chinese thought), than in its engaging style, novel ideas, and rigorous argumentation, which can serve as a model for future work in Chinese philosophy. Hansen takes Chinese philosophy seriously as philosophy. For anyone tired of the superficial summaries or scholastic commentaries that so often characterize this field, Hansen's book will be a memorable and welcome change.

MICHAEL R. MARTIN

University of Hong Kong

The Legacy of Wittgenstein. ANTHONY KENNY. New York: Basil Blackwell, 1984. xv, 150 p. $24.95.

The Legacy of Wittgenstein collects ten of Anthony Kenny's previously published essays, spanning a period of twenty years, which have been revised where necessary to free them from their original contexts. Each essay is associated in one way or another with Wittgenstein's work: four are devoted to the examination of Wittgenstein's own ideas (his early philosophy of mind in the *Tractatus,* his later misrepresentations of his earlier views, the relation of the "Big Typescript" to the published *Philosophical Grammar,* and Wittgenstein's view of the nature of philosophy in the central period of his life); three relate Wittgenstein's work to major philosophers of the past (Aquinas, Descartes, and Russell); and three expose philosophical confusions, diagnosed by Wittgenstein, present in the writings of Teilhard de Chardin, R. L. Gregory, and Noam Chomsky. Anyone who is interested in Wittgenstein's work will find the essays interesting and profit from studying them: they are clear, penetrating, and closely argued. I will comment on one essay, in which Kenny does less than justice to Wittgenstein and, I am sure, to his own knowledge of Wittgenstein's thought.

In "Intentionality: Aquinas and Wittgenstein," Kenny proposes to construct a viable account of the nature of thought by wedding Aquinas's theory of the immaterial existence of forms in the mind to Wittgenstein's solution to "the problem of the harmony between world and thought." But the quotations from Wittgenstein's writings that Kenny selects and the use he makes of them do not adequately represent the scope of this problem or Wittgenstein's solution to it:

0022-362X/87/8401/0042/$00.50 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

they narrow the scope to a question that concerns only thoughts about individuals. And even within this restricted scope Wittgenstein's solution remains unclear. When the real scope of Wittgenstein's problem is restored to it, there is no room for Aquinas's theory as a supplement to Wittgenstein's solution.

Wittgenstein's treatment of the topic of the agreement or harmony of thought and reality has two principal aspects. The first seems trivial when considered in isolation. My thought, intention, expectation, wish, hope, or command has a content that may or may not agree with reality: a thought can be true or false, an intention carried out or not acted upon, an expectation or a hope fulfilled or unfulfilled, a wish satisfied or unsatisfied, and a command obeyed or not. What is the connection between my thought, intention, expectation, wish, hope, or command and reality when reality agrees with it? To take the case of expectation: if my expectation is described as the expectation that *p*, the state of affairs that fulfills this expectation is described as the state of affairs that *p*. This is the sense of Wittgenstein's remark, "It is in language that an expectation and its fulfilment make contact," and, generalizing, of his remark, "Like everything metaphysical the harmony between thought and reality is to be found in the grammar of the language." The item's content and its fulfillment, satisfaction, or whatever, are rendered into language in the same words. And a description of an expectation in terms of what is expected is an *internal* description of it, as it would not be if were like a description of a state of hunger in terms of what might possibly remove it. Furthermore, it is essential that there should be such an internal connection between thought and reality. For if, *per impossibile,* there were not, there would be no sense to the idea of the content of a thought.

Now this first aspect derives its significance from the second. For —to continue with the case of expectation—although my expectation and what happens when it is fulfilled are described in the same terms, the agreement between them is not a consequence of the fact that what happens was present in my expectation. That would not be possible. But if the agreement consists in the fact that if my expectation has a certain content, it follows that its fulfillment is described by the words that specify that content, the question arises, In virtue of what is it true that my expectation does have that content? And a similar question can be raised about my thought, intention, wish, hope, or command. The second aspect of Wittgenstein's consideration of the problem of the harmony between thought and reality concerns the relationship between what happens when I think a certain thought, form a certain expectation, or whatever, and my

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

thinking a thought with that content or forming that expectation. His main preoccupation was with the idea that something other than and essentially different from a mere sign must be present to my mind when I think, expect, intend, wish, hope, or command in order for the requisite internal relation between me and reality to obtain. This illusion is encouraged by the undeniable fact that whatever signs do go through my mind, they could be understood in different ways. But—Wittgenstein's argument is familiar—nothing else that went through my mind would itself establish an internal relation of the required kind between me and reality. Any supposed intermediary in my mind connecting the expression of my thought in signs with the reality with which my thought is concerned will stand in no closer relation with reality than the signs themselves: a "method of projection" will always be needed to effect the transition from what is in the mind to the reality it represents. Now this realization can make it appear that, no matter what happens in my mind at a certain time, I cannot then be connected with reality in the way that is required if I am to think a thought. But the truth is that there are facts about me other than facts about occurrences in my mind, and these can make it correct to describe me as thinking a thought with a certain content —thereby making it possible for there to be the required harmony (or lack of harmony) with reality.

We can now consider two of Wittgenstein's illustrations which Kenny singles out. If what is in my mind is an image, what makes it true that my image is an image *of John* is not the nature of the image as specified by spatial/colored terms and the similarity of the image to John. And if what passes before my mind is a series of words, what makes it true that I have at this moment thought *of John* is not any intrinsic features of the signs that go through my mind when I think of him. What occurs in my mind is describable in the way it is—as an image or the expression of a thought *of John*—in virtue of facts other than the state of my consciousness, and, in particular, in virtue of my mastery of a language, my past and present situation, and what I *would* do. It is for this reason that, as Wittgenstein emphasized, if God had looked into my mind he would not have been able to see there whom I was visualizing or thinking of. But it would be wrong to focus upon just these examples and to construe Wittgenstein's problem of the harmony of thought and world as concerned only with the question, What makes an image or a sentence or picture one that is of or about a certain individual *N*? For, as I have indicated, the problem Wittgenstein is dealing with is completely general and concerns equally thoughts that are not about individuals. If God had looked into my mind when I had such a thought, he would have been

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

able to see there not the content of my thought, but only any signs that constituted the vehicle of my thought: unless the signs are understood in accordance with a particular method of projection they do not disclose the content of my thought. Kenny's idea is that the answers of Aquinas and Wittgenstein to the question, What makes a thought of *X* to be a thought of *X*? can be seen as complementary, since Aquinas's concerns the *sense* of our thoughts and Wittgenstein's their *reference*. But this omits most of Wittgenstein's contribution to the topic, and when this is accurately represented it needs no supplementation of the kind Kenny proposes. What it does need is something I have not been able to give it in this review—detailed exposition and examination.

MALCOLM BUDD

University College, London

Probabilistic Metaphysics. PATRICK SUPPES. Oxford: Blackwell, 1984. xii, 251 p. Cloth $24.95, paper $15.95.

This is a very general work, containing discussions of logical completeness and rationality, as well as a number of chapters on causality, determinism, and probability. I believe the chapter with the most original substance is the long chapter on probabilistic grammars. Since that is the area in which I have least competence, I shall leave that chapter for other reviewers and concentrate on the topics most germane to the title: the topic of causality and determinism and the topic of probability and uncertainty.

CAUSALITY

We have here a new statement of Peirce's metaphysical doctrine of tychism: the doctrine that there is an inherent and ineradicable chance element in the course of the universe. Suppes claims to have "widened the arena" (11), at least by incorporating a consideration of probabilistic grammars. The key ideas—at least those which justify the title of the volume—are that the fundamental laws of nature are essentially probabilistic and that causality itself is probabilistic, so that there is no conflict between "randomness in nature" and "the existence of valid causal laws."

Let us look more closely at the question of randomness in nature. Suppes contrasts it with determinism—for example, the fact that in particle mechanics the trajectories of a set of particles are determined by their positions and velocities at some one instant and by the

0022-362X/87/8401/0045$00.50 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

forces acting on them. "In contrast, . . . [suppose] that we are going to flip two coins a thousand times . . . We begin with the same initial conditions . . . [and] ask ourselves then, what is the probability that the two coins will have exactly the same sequences of heads and tails? The answer is easy to compute. First, we can ask what is the probability on each trial that the two coins will either both show heads or both show tails? The answer is of course 1/2" (19).

"Of course"; but many writers would say that this contradicts the "sameness" of the initial conditions, since coins are not generally regarded as quantum-mechanical devices. Suppes—and this is typical of the style of argument—nowhere addresses this issue of the grounds on which we might or might not want to say that the "initial conditions" were "the same"—clearly an important issue in this context.

More generally, Suppes sees the introduction of randomness into our physical theories of the universe taking place in several stages over the past few hundred years. First there is the "recognition that measurements . . . are subject to errors" (20). Then there are systems subject to random external perturbations. And there are the cases (hydrodynamics, statistical mechanics) where, despite the conviction that the underlying process follows classical deterministic laws, the only feasible way of attempting to deal with these phenomena is through stochastic equations. Finally, of course, we have quantum mechanics to fall back on. It is only the last that gives serious support to Suppes' thesis.

This point may be worth expanding. Measurement: the *system* consists of an exactly and precisely determined magnitude. Our measurements of it yield a population of values. To speak of *errors* of measurement already presupposes the contrast between the population of observed values and the underlying true value. In order to see indeterminism here, we would have to reject this contrast: to say that the *system* is to consist of a population of values. But then we need no notion of "error" at all. Suppes suggests that the "classical response" to errors of measurement is to claim that "errors of measurement can in principle be eliminated." I have never heard this response; and it certainly seems quite irrelevant to the ordinary analysis, according to which the system is deterministic and it is measurement that introduces error. Suppes himself suggests that one should question the assumption of an underlying real value; but he does not show us how to do it.

Similarly in the second case: We do not classically think of the system as responding indeterministically to influences from outside just because we can't predict the effects of the perturbing influences.

CC-0. In Public Domain. Panini Kanya Maha Vidyalaya Collection. Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

And in the third case it is sometimes in virtue of underlying deterministic assumptions that we can derive the stochastic equations characterizing a certain process. (Though of course there are also occasions when we can get the stochastic equations only "empirically".)

Quantum mechanics, as currently construed, does indeed provide evidence of "randomness in nature"; but Suppes evidently wants to go beyond that: "It is not just the quantum world that has an essential random component" (29). It is unclear how. Suppes refers to the "fantasies" of Descartes, he calls the deterministic idealizations of classical physics "among the most misleading aberrations of philosophical thought since the pre-Socratics," and then, remarking on "the endless variety of natural phenomena," he suddenly concludes that the "evidence supports the thesis that random or probabilistic phenomena are found in nature and not simply in our lack of knowledge."

Suppes offers a characterization of causality that is consistent with the notion that nature is probabilistic. This follows the lines laid down earlier.[1] The procedure is to define: B is a *prima facie cause* of A if B occurs earlier, and $\text{Prob}(A/B) > \text{Prob}(A)$; B is a *spurious cause* of A if B is a prima facie cause of A and if there is a "partition" C_i of prior events such that, for all i, C_i is earlier than B and $\text{Prob}(A/B \,\&\, C_i) = \text{Prob}(A/C_i)$. B is a *genuine cause,* otherwise.

This definition is the occasion for one of the examples of polemics that occur (I am tempted to say "at random") throughout the volume: "The history of human folly is replete with belief in spurious causes" (51). No doubt. Suppes refers to astrology. I suppose almost all of us would agree that astrology is not good science—but Suppes does not even say what considerations *would* show that spurious causation is the problem, much less give us any examples of these considerations. In fact, it would not be easy. For instance, Ptolemy is cited (there are many exotic citations in the book) as asserting that the citizens of countries in closer familiarity with Aries (such as Britain) tend to be more bestial. This is alleged to be an instance of spurious causation.

So we should *first* show that familiarity with Aries increases the probability of bestiality—i.e., that it is a prima facie cause of bestiality. (Suppes does not suggest that Ptolemy did this, or even that he could have.) *Then* he should show us the particular partition C_i of events, all earlier than familiarity with Aries, such that, within each class of the partition, familiarity with Aries does not alter the probability of bestiality.

[1] *A Probabilistic Theory of Causality* (Amsterdam: North Holland, 1970).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

UNCERTAINTY

Leaving metaphysical questions of causality and determinism to one side, we may consider Suppes' treatment of uncertainty. He is concerned to find some compromise between "objective" and "subjective" points of view (98). The illustrative example is between (1) the weight of the Chevy is between 2000 and 3000 pounds, and (2) I believe it weighs 2800 pounds. The latter is "subjective"; the former is "objective." This is clearly absurd. One statement concerns weights; the other concerns beliefs. Both are most naturally construed as objective. Both *can* be construed as "subjective"—namely when what is being reported is *your* opinion concerning my belief or *my* opinion concerning the weight. [Note that the truth of (2) is completely independent of the weight of a Chevy.] New evidence about the weight of the car may correct (1); only psychological evidence can correct (2).

In general the discussion of probability is more confusing than enlightening. There is the cryptic remark, for example, that "there is a large literature by advocates of a relative frequency theory of probability about how to make inferences from relative frequencies to single cases" (61). In general, a view that identifies probabilities with relative frequencies can make no sense of probabilities applied to single cases. To the best of my knowledge there is nothing in the statistical literature (at least for the last thirty years) and very little in the philosophical literature, concerning "direct inference"—the inference that leads from frequencies (or propensities) to "single cases."

Suppes takes Laplace severely to task early on for thinking that probability has to do with knowledge and ignorance, rather than with randomness in nature (Chapter Two), but then unblushingly relativizes his own probability to knowledge: "Of course, if I learn something new, the probability can change" (61).

Finally, in the last chapter, on rationality, probability again comes in for discussion. Suppes mentions the possibility of assigning upper and lower probabilities, attributing the general idea to Thomas Bayes, but not mentioning any of those (I. J. Good, C. A. B. Smith, and others) who have in the past thirty or forty years attempted to put flesh on the bare idea. He offers the idea of using standard events—coin tossings, and the like—to calibrate our degrees of belief in other matters, without mentioning Glenn Shafer, who has developed this idea in connection with his own representation of evidential import.

A very general question is raised by this last chapter. It is quite clear that our knowledge is infected with uncertainty—that probabil-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ity is central to decision theory and to belief theory. Given that this is the case, has not the argument for locating probability and randomness in the world, rather than in our view of the world, been undermined? Do we really need both? Quantum mechanics teaches that we need indeterminism in the world; but that is an argument against causality, and not an argument in favor of an unspecifiable metaphysical notion of randomness. There are frequencies and, we now have reason to believe, propensities in the world. There are foolish opinions as well as sound ones. It is not the message that Suppes is purveying, but the message I get from the material of this book is that we should seek the rational or inferential connection between those frequencies and propensities, and rational or justified opinions.

HENRY E. KYBURG, JR.

University of Rochester

Intending and Acting: Toward a Naturalized Action Theory. MYLES BRAND. Cambridge: Bradford Books/MIT Press, 1984. xiii, 296 p. $27.50.

The connection between interpreting the world and changing it—between cognition and conation—in one form or another has long been of interest to philosophers. It is no news, however, that the theory of knowledge has been transformed by the cognitive revolution, and it is remarkable, then, that the closely allied theory of action has remained untouched by these developments and almost entirely aloof from the goings on in artificial intelligence, psychology, and neuroscience. There is some irony in this fact, since it was a work specifically concerned with human action which was in the vanguard of the cognitive revolution.[1]

There is reason, therefore, to welcome Brand's boldly proclaimed goal of ushering in the next and naturalized stage of philosophical action theory, after what he describes as the stagnation and waning interest of the 1970s. On the other hand, there is corresponding reason to lament what I believe to be Brand's own conspicuous failure to achieve these goals.

Proportional to the expectations raised by Brand's rhetoric is the reader's inevitable exasperation on reaching the seventh chapter (of

[1] George A. Miller, Eugene Galanter, and Karl H. Pribram, *Plans and the Structure of Behaviour* (New York: Holt, Rinehart & Winston, 1960).

0022-362X/87/8401/0049/$00.50 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

nine) before there is any hint of a concern with science and its possible bearing on philosophical theories of action. Until this point there are 170 pages of just the kind of arid scholasticism which Brand had promised to end. Worse still, the bulk of these chapters is only very remotely related to the problems of action at all. Even where he is dealing with actions, however, Brand demands a great deal of patience of his reader through his *penchant* for spurious formalism, which serves only to detract from the readability of the book. For example:

> *If* (i) S has an action plan P such that his $A_i$ing and his A_{i+1}ing are contained in P but his A^*ing is not contained in P, (ii) S's A^*ing occurs between S's $A_i$ing and S's A_{i+1}ing or during S's $A_i$ing or S's A_{i+1}ing and (iii) S's A^*ing is not dependent on either S's $A_i$ing or S's A_{i+1}ing, or conversely, *then* S's $A_i$ing and S's A_{i+1}ing are intentional actions (29).

The essence of Brand's concern in the book is to account for what he takes to be possibly "*the* fundamental problem" in philosophical action theory, namely, "to adequately specify the nature of the proximate cause of action" (33). For Brand, this fundamental problem about action concerns the nature of the mental antecedent such that only events of this type initiate action. He says

> I take it to be a deep insight of folk psychology that the initiating mental event has both cognitive and conative features. Without conative features, action could not begin (45).

Thus, he explains

> My objective is to show how folk psychological action theoretic concepts can be transformed into scientific psychological ones (191).

In short, Brand is concerned to seek a scientific foundation for the will.

Although Brand's search for the will has the considerable virtue of making it a scientific enterprise rather than a metaphysical one, I will suggest that this virtue is greatly diminished by his eccentric view of science.

Brand begins encouragingly enough by claiming (x) that he will argue that philosophical and scientific theories are continuous. This is the now familiar view of W. V. Quine's "Epistemology Naturalized,"[2] and Brand reassuringly emphasizes that his approach is in the spirit of "some recent developments in epistemology, the philosophy of mind, the philosophy of language and value theory" (x), citing Jerry Fodor and Daniel Dennett as exemplars. However, Brand's

[2] In *Ontological Relativity and Other Essays* (New York: Columbia, 1969).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

conception of this continuity has little in common with the tradition he claims to represent. Instead of seeing a single enterprise with shared substance, goals, and principles, Brand takes science to be significantly distinct from philosophy and folk psychology (the latter two not being clearly distinguished). For Brand, folk psychology has a privileged status whose all but infallible dictates are to serve as the Cartesian foundation upon which to construct science. He explains, "my goal is to build from the bottom up" (xi) and "philosophical action theory is the foundation on which I build" (xi). A further methodological principle enjoins us to "point directions on empirical issues" from the allegedly more secure foundation of philosophy and common sense, and it is clear that Brand does not see himself aboard the Quine-Neurath boat. Brand's view that there is such a privileged vantage point outside science is justified on the grounds of the corrigibility of science (xi), which is again curiously Cartesian for a contemporary naturalistic philosopher:

> Since the models used in these disciplines are evolving and, sometimes, changing radically, it would be a mistake to inexorably tie an account of action to specific models (xi).

Accordingly, Brand disavows any concern to engage in actual scientific speculation or adjudicate between competing theories, which he says would not be appropriate since "this is not an essay in empirical psychology" (174). However, like Descartes before him, Brand not only predictably makes "sidelong glances" into science but indulges freely in substantive criticism and speculation.

For Brand, the continuity between science and philosophy involves the possibility of some kind of metamorphosis in which folk-psychological concepts become scientific through "progressive transformation." In this way, for Brand (46), philosophical considerations provide constraints on psychological theorizing, and he speaks of the need for scientific psychology to be "tempered by folk psychology" (157). Accordingly, when Brand does consider certain psychological models or theories, it is for the express purpose of coopting them to philosophical purposes for which they were, of course, never intended. Brand insists that it is "an essential ingredient in the folk psychological picture of action that it is initiated by a mental event" (247), and it is on such an a priori basis that he complains that scientific theories are inadequate when they appear to be "unlikely candidates for the basis of a transformation of the conative feature of immediate intention" (247)—just as medieval anatomists would blame a cadaver when it failed to confirm Galenic principles. In a typical example, Brand considers Hullian Drive Theory and con-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

cludes that it is unsatisfactory through its failure to reflect distinctions found in folk psychology (248). Brand explains:

> No progress can be made in clarifying and articulating the conative feature of intention in terms of drive, since that construct fails to serve a scientific explanatory role. It is as if an articulation of the commonsense notions of force and motion were attempted to be gained by transforming them into constructs within Aristotelian physical theory (249).

This latter analogy is bizarre, though symptomatic, since it rests on the wisdom of hindsight in correctly noting that we would not employ a theory known to be wrong. The relevant point, however, is, surely, that we cannot do better than endorse our best theories. Their failure to accommodate some notion is ordinarily taken to be grounds for rejecting that notion. Brand's entire book, with repeated examples of this kind, can be read as making a strong case *against* the scientific respectability of conative features of action. Brand seeks to superimpose a philosophical map on a foreign terrain and concludes from the mismatch that the terrain is at fault. Or, like the English-speaking tourist in Europe, Brand complains of all the bloody foreigners. Less figuratively, Brand only grudgingly concedes (168) that folk psychology might in principle need to be abandoned in the light of systematic scientific research. His unexplained commitment to the sanctity of our folk theories, it must be noted, is quite different from the mere confidence in their over-all reliability to be seen in Fodor, for example. Of some current work in the spirit of William James's ideomotor theory Brand says:

> But these theories, no matter their appeal, do not fully explain human action. They do not provide a transformation of the folk psychological concept of action because they leave out its essential property of being caused by a conative event . . . scientific psychology is not free to develop any arbitrary conceptual scheme; it is constrained by the conceptual base of folk psychology (239).

> Action theory is naturalized by bringing its conceptual basis closer to that of scientific psychological theories, not by eliminating it in favor of some particular scientific psychological theory (169).

To be sure, there are plenty of cases in which folk concepts end up finding a place in systematic science, but this is more through good fortune than any methodological imperative. In Fodor's apt slogan, there is no science of what happens on Tuesdays.

Among Brand's choice of theories, old and new, onto which he might graft conative features, noteworthy for its glaring omission is the vast literature of the neurosciences, where we have, as it happens,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the richest and most detailed understanding of motor control—and, of course, in particular, the causal antecedents of action.

Besides his practice, what Brand offers as explicit methodological principle only serves to confirm the diagnosis I have given. Tucked away inconspicuously, as an unlikely subsection of chapter six on intending and believing, is an account of the relation of folk psychology to systematic science. Brand begins by stating that both folk psychology and scientific psychology can be modeled by the formalism of ordered pairs (160) which is, however, utterly flatulent here since it is unilluminating and remains unexplained, playing no further role in the exposition at all. Brand proceeds to introduce a notion of "conceptual psychology" (163) which he takes to be certain claims shared by folk psychology and scientific psychology additional to their empirical claims:

> . . . on the whole there is a recognizable tripartite distinction between folk psychological generalizations, scientific generalizations, and the conceptual claims that support and define both types of generalizations (163).

In case this is not transparently clear, Brand explains "conceptual psychology" by saying that it can be modeled by a three-dimensional space in which the x-coordinate is the degree of theoretical relevance, the y-coordinate the degree of empirical support, and the z-coordinate is time, the statements of conceptual psychology being construed as points in this conceptual space. In case this is still obscure, Brand offers examples of "conceptual psychology" such as the following: "there is a linguistic code in which persons think," "mental events have syntactical (formal) properties" (165), "some thinking is imagistic" (163). Even if one didn't have reservations about analytic and synthetic, however, these statements would appear to be straightforward paradigms of empirical claims. By this time one might justly conclude that any unclarity and confusion is Brand's fault.

Brand's method recommends "the change of conceptual claims underlying folk psychology into ones that underlie scientific psychology," calling the process "progressive transformation," which, in an idealized form, he labels "the Method of Case Studies." To the extent that any of this is clear, construed descriptively, it is a dubious account of scientific practice, and, construed normatively, it suggests the obvious question: Why bother? Without the galenist's prior convictions concerning the *rete mirabile*, this elaborate and contrived algorithm for scientific progress has little to recommend it.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

In the end, even on his own account, Brand's project is unsuccessful as he laments:

> Unfortunately the situation for conative features is less happy [than for cognitive ones]. There is presently no tenable scientific theory into which this folk psychological concept can be transformed (237).

Using a strategem popular with purveyors of the paranormal, Brand takes his failure to reflect adversely on science rather than conation, and, in the absence of positive evidence, he resorts to the entirely vacuous claim that future science will vindicate him. He says "a complete transformation of conation must await future scientific developments" (237), thereby unwittingly placing conation in the same category as all the other things which presently have nothing to recommend them.

In these heady days of cognitive science (a term not used by Brand), the a priorism of his project appears as a kind of philosophical atavism, a regressive program which seeks some psychological equivalent of the pineal gland as a repository for the will—though probably in this case a *will*-o'-the-wisp.

PETER SLEZAK

University of New South Wales

NOTES AND NEWS

The London School of Economics and Political Science is pleased to announce that the first Lakatos Award has been given to both Hartry H. Field (Southern California), for his book *Science without Numbers*, and Bas C. van Fraassen (Princeton), for his book *The Scientific Image*. The Award is made for an outstanding contribution to the philosophy of science, and has been generously endowed by The Latsis Foundation in memory of Imre Lakatos. It will be awarded annually for at least ten years. Recipients are expected to deliver at the School a public lecture of interest to a general audience.

The closing date for nominations for the second Lakatos Award is April 15, 1987. The Award will be conferred on the basis of a book published in English during the last ten years and judged to have made an outstanding contribution to the philosophy of science. At least three nominators should give their grounds for nomination, and provide the nominee's age, since preference may be given to younger scholars. Nominations should be marked "Lakatos Award" and addressed to: The Secretary, The London School of Economics and Political Science, Houghton Street, London WC2A 2AE, England.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The Department of Philosophy of the University of York, under the auspices of The Royal Institute of Philosophy, is pleased to announce a conference on Philosophy and Medical Welfare, to be held September 11–13, 1987. Speakers include Martin Hollis (East Anglia), John Harris (Manchester), Michael Lockwood (Oxford). Further information may be obtained from the Philosophy Conference Secretary, Department of Philosophy, University of York, Heslington, York Y01 5DD, England.

The Joint Guelph/McMaster Graduate Department of Philosophy and the Ethics Research Group at the University of Guelph are pleased to sponsor a conference on Justification in Ethics, to be held at Guelph, May 22–24, 1987. Topics addressed will include: Justifying Morality, Justification and the Environment, Justification and Foundations, Justification and Diversity, Justification and Coherence, Agreement in Ethical Theory. Speakers will be: Leslie Armour (Ottawa), Ted Bond (Queen's), Richard Campbell (Dalhousie), David Copp (Chicago), Danny Goldstick (Toronto), Trudy Govier (Calgary), Marsha Hanen (Calgary), Harold Johnson (Western), David MacNiven (York), Peter Miller (Winnipeg), Jan Narveson (Waterloo), Harry Nielsen (Windsor), Kai Nielsen (Calgary), Joe Novak (Waterloo), Ken Rankin (Victoria), Lionel Rubinoff (Trent), and Roger Shiner (Alberta). Further information may be obtained from: Doug Odegard, Philosophy, University of Guelph, Guelph, Ontario N1G 2W1.

The Francis Bacon Library of Claremont, California announces that the 14th Annual Bacon Birthday Celebration will be held January 22, 1987. The Celebration is sponsored by the Library in cooperation with the Claremont Graduate School. Peter Urbach (London School of Economics) will speak on Bacon's Modern Philosophy of Science. A reception will follow at the Library, where rare books and manuscripts will be displayed. Further information may be obtained from The Francis Bacon Library, Office of the Director, 655 Dartmouth Avenue, Claremont, Calif. 91711.

The Long Island Philosophical Society invites the submission of papers for its conference, Person and Community: The Philosophy of Max Scheler, to be held at SUNY/Stonybrook on April 23/24, 1987. Papers on all aspects of Scheler's work are welcome. Especially encouraged are contributions by graduate students, and a special prize will be awarded by the Society for the most significant contribution to Scheler studies made by a graduate student. More information may be obtained from: Eugene Kelly, Chairman, Long Island Philosophical Society, New York Institute of Technology, Central Islip Campus 66325, NY 11722, or from: Kenneth Stikkers, Department of Philosophy, Seattle University, Wash. 98112.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

MEMO TO AUTHORS

The editors cannot afford to send to the printer manuscripts that are difficult for the compositor to read or footnotes that need to be completed in proof. They ask that manuscripts submitted be prepared according to the following guidelines:

1. All material, including the quotations and footnotes—especially footnotes—must be typed, double-spaced, on one side of the paper, with ample margins left and right, top and bottom.
2. No bibliographies will be published; bibliographical material should be put into the footnotes.
3. Footnotes should be complete, according to *Journal* style (see any issue). Journal citations should include name of journal in full, volume number, issue number, month or season, year, and pages. Book references should include publisher, city of publication, and date. Footnotes should be gathered at the end of the article.
4. Quotations should be checked by the author; capitalization, italics, spelling, etc., should match the original. Corrections should not have to be made (expensively) in proof.
5. If there are diagrams, they should be provided camera-ready by the author. If there are special symbols, they should be widely spaced and clearly marked; if there are many, a list should be provided. Standard logic symbols are available, and letters of the greek and script alphabets; no boldface.
6. Computer copy can be extremely difficult to read; only letter-quality copy is acceptable. It should be double-spaced and not justified. The print should be dark enough and the letters widely enough spaced so that both letters and punctuation marks can be read easily.
7. Finally, photo copies on shiny paper are not acceptable. Ribbon copy is much preferred, but clear copies on bond paper are acceptable if necessary.

The editors are willing to read and evaluate manuscripts that do not meet these standards, provided they are genuinely legible and not unreasonably long (over 15000 words). They ask for revision of excessively long manuscripts, and they require retyping of accepted manuscripts, if necessary, before scheduling for publication.

The *Journal of Philosophy* is published monthly, in issues that (ordinarily) contain only 56 pages of copy; this limits the length of papers the editors can normally consider, to no more than 7,500 words, or eighteen JOURNAL pages. Manuscripts very much in excess of these limits will be returned unread.

The *Journal of Philosophy* does not publish papers that are primarily historical or expository; the editors urge that such manuscripts be submitted elsewhere.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

JOHNSONIAN PRIZE IN PHILOSOPHY

The editors are pleased to announce that the winner of the Johnsonian Prize in Philosophy for 1986 is Clyde L. Hardin, for his book *Color for Philosophers,* sponsored by the Department of Philosophy at Syracuse University, where Professor Hardin teaches. *Color for Philosophers* will be published in 1987 by Hackett Publishing Company, the donor of the prize.

Honorable Mention goes this year to Dionysios Anapolitanos, for his *Leibniz: Representation, Continuity, and the Spatiotemporal,* sponsored by the Department of Philosophy of the University of Pittsburgh. Dr. Anapolitanos is now teaching at the University of Athens, Greece.

Previous Johnsonian prize winners have been: 1980, Daniel Bonevac, *Reduction in the Abstract Sciences* (Indianapolis: Hackett, 1981); 1981, Igal Kvart, *A Theory of Counterfactuals* (1986); 1982, Moreland Perkins, *Sensing the World* (1983); 1984, Robert J. Dunn, *The Possibility of Weakness of Will* (1986); and 1985, Harry van der Linden, *Kantian Ethics and Socialism* (forthcoming in 1987). Also, honorable mention in 1981 went to Charles B. Guignon, *Heidegger and the Problem of Knowledge* (1983).

The Johnsonian Prize, which consists of one thousand dollars and guaranteed publication of the prize-winning book, will be offered again in 1987. The prize is designed to encourage the work of unpublished philosophers; authors are eligible who have published no book-length work. Books are selected from among manuscripts sponsored by departments of philosophy. Authors and departments who wish to submit manuscripts should observe the following guidelines:

1. Any graduate school of philosophy that confers the Ph.D. degree may submit one manuscript per year. It is not required that the work sponsored by a department be written by either a graduate or a faculty member of that department; it may, but need not be a revised version of a departmental dissertation. The closing date for this year's prize is August 1, 1987.

2. To qualify, the manuscript must be: (a) of book length (at least 40,000 words) and in proper form for publication as a book; (b) accompanied by an abstract of up to 2000 words; (c) written by someone who has not yet published any book-length work; (d) accompanied by two professional evaluations and a letter from the sponsoring department (the editors are depending on the departments to screen out all but the most important and original work); (e) uncommitted to a publisher. Manuscripts will be returned to the authors only if accompanied by appropriate postage.

3. Graduate schools who wish to participate should notify the JOURNAL in advance of submitting a manuscript for consideration, and, if possible, send evaluations and abstracts in advance.

4. Preference will be given to the work of younger philosophers, especially those who have not yet been granted tenure.

Announcement of the seventh award will appear in the January 1988 issue of this JOURNAL; some manuscripts may be given honorable mention. Publication of the prize-winning book will be guaranteed by Hackett Publishing Company after the author has made what revisions are deemed necessary by the editors, if possible within twelve months of the award. The sponsors hope that the prize-winning books for both 1985 and 1986 will be on display at the APA Eastern Division meetings in 1987.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

New from Princeton

Now Available in Paperback

Foundations of Space-Time Theories

Relativistic Physics and Philosophy of Science

Michael Friedman

Winner of the Franklin J. Matchette Prize

"...any serious student of this difficult subject should read this book, which is an excellent introduction to the philosophies of space and time. 'Foundations of Space-Time Theories' is the result of the author's intellectual struggle to clarify the foundations of relativity theory for himself, and it will deeply influence all future thinking about the philosophy of space and time." —*Heinz R. Pagels, The New York Times Book Review*

P: $14.50. C:$37.00

Nietzsche in Russia

Edited by *Bernice Glatzer Rosenthal*

Here presented are eighteen essays and a bibliography on the Russians' fascination with Nietzsche from 1890 to 1917, with glimpses of developments in the Soviet period. In this interdisciplinary volume scholars document Nietzsche's enormous impact on Russian cultural and political life—an influence that has until now been ignored, understated, or denied in both the United States and the Soviet Union. Illustrated.

LPE:$19.95. C:$50.00

At your bookstore or

Princeton University Press

41 William Street, Princeton, NJ 08540

Theory and Decision

An International Journal for Philosophy and Methodology of the Social Sciences

Editor-in-Chief: W. LEINFELLNER

Some Recent Articles

Timo Airaksinen: Coercion, Deterrence, and Authority. **James C. Gaa:** The Stability of Bargains Behind the Veil of Ignorance. **Paul Grout:** Good Approximations to Maximum Allocations. **Yek-Kwang Ng:** Interpersonal Level Comparability Implies Comparability of Utility Differences. **Leigh Tesfatsion:** Games, Goals, and Bounded Rationality. **Lanning Sowden:** Rule Utilitarianism, Rational Decision and Obligations. **Paul Weirich:** The St. Petersburg Gamble and Risk.

Subscription Information

1986, Volume 20–21 (6 issues) ISSN 0040–5833

Institutional rate: Dfl. 372.00/US$ 132.00 incl. postage/handling

Private rate: Dfl. 144.00/US$ 52.00 incl. postage/handling

Private subscriptions should be sent direct to the publishers.

D. REIDEL PUBLISHING COMPANY

A Member of the Kluwer Academic Publishers Group

P.O. Box 17, 3300 AA Dordrecht, The Netherlands

190 Old Derby Street, Hingham MA 02043, U.S.A.

Falcon House, Queen Square, Lancaster LA1 1RN, U.K.

CC-0. In Public Domain. Sri Muthulakshmi Kaagati Collection, Madurai

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIV, NO. 2, FEBRUARY 1987

HUMAN BEINGS*

MANY of us hope for a kind of philosophy that is precise without thereby desiccating its object, so that the results of a philosophical investigation could answer a question still worth asking. So it is with the question: What is it that we are? But much of contemporary philosophical theorizing about personal identity conforms to what could easily become a desiccating analytic paradigm. Cases imaginary and real are produced. Competing accounts of the necessary and sufficient conditions for personal identity are then evaluated simply in accord with how well they jibe with intuitions wrung from these cases.

Compare Gettierology, the search for the analysis of knowledge by means of conjecturing necessary and sufficient conditions for knowledge, followed by confirmation or refutation by appeal to intuitions about cases.[1] Here reliance on the method of cases has a straightforward, if somewhat wooden, justification. We are interested in articulating our concept of knowledge and in charting its relations to other concepts, such as the concepts of believing a proposition, of that proposition's being true, of one's being justified in believing the proposition, of one's employing a reliable method in coming to believe the proposition, and so on. Masters of these concepts by and large agree about when they are to be applied, and so the method of cases appears well adapted to investigating the relations among them.

* In writing this paper I have been helped by conversations with Rogers Albritton, Paul Benacerraf, Philippa Foot, David Kaplan, Saul Kripke, David Lewis, Derek Parfit, Warren Quinn, and Michael Smith. My indebtedness to the works of Bernard Williams on personal identity should be obvious.

[1] For an extensive survey of this material, see Robert K. Shope, *The Analysis of Knowing* (Princeton, N.J.: University Press, 1983). I do not mean to suggest that the method of cases is entirely unproblematic in epistemology, only that it is even more problematic when it comes to personal identity.

0022-362X/87/8402/0055$02.50 CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar © 1987 The Journal of Philosophy, Inc.

Digitized by Arya Samaj Foundation Chennai and eGangotri

But what relations among concepts are being charted when we develop a systematic account of the conditions under which a person survives, continues on, or persists through a period of time? No doubt there is some innocuous sense in which we are employing some nonindividual concept when we judge that Anthony Blunt is the same person as the notorious Maurice or that Orcutt is not the same person as the most cultured spy. No doubt there is some relation between this concept of being the same person and the concepts of mental continuity, of physical continuity, and of the various patterns of causal dependence which could be responsible for such continuities. However, it ought to be controversial that the relation here is the same as the relation between knowledge and, say, true belief. Believing correctly that *p* is arguably part of what it means to have knowledge that *p*, whereas the holding of patterns of mental and physical continuity and dependence across a stretch of time seems to be the sort of evidence we might ideally rely upon in justifying our claim that a person survived or continued on through the stretch in question. Why should we think that we could analyze what it means to say that someone has survived in terms of the concepts of the sorts of relations that are *evidence* for survival?

The method of theorizing about personal identity solely or mainly by appeal to our intuitive reactions to puzzle cases which exhibit all sorts of variations in kind and degree of continuity and dependence would be justified if two requirements were satisfied. First, our grasp of the concept of being the same person should be able to be correctly represented as a grasp of necessary and sufficient conditions for the application of the predicate "is the same person," conditions that could be cast in terms of statements about continuity and dependence, statements not themselves to be explained in terms of statements about personal identity. Call this the *reductionist requirement*. Secondly, our intuitive reactions to the puzzle cases should be able to be taken as manifestations of our grasp of those necessary and sufficient conditions, and not as overgeneralizations from the everyday run of cases or manifestations of a particular conception of people, be it a religious conception (e.g., that people are reincarnatable souls) or some more inchoate secular counterpart. But then it should be evident that this second requirement generates a difficulty for the ideology behind the method of cases. Given the enormous variety of apparently conceptually coherent *conceptions* of people which have been entertained, we must assume that our common *concept* of people, if there is such a thing, is quite unspecific; so the topic of personal identity will be the dry articulation of a vague generality compatible with all the more interesting and specific conceptions that have guided practical life.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

This difficulty not having been taken seriously, there is a dominant—and as it turns out, fairly specific—philosophical view about people which is defended in large part by the method of cases. The dominant view has it that we are minds—perhaps essentially embodied minds, although no *particular* body or brain is such that its survival is essential to ours. Put in more familiar terms, the dominant view is what we might call, following Derek Parfit, *wide psychological reductionism.*[2] *Psychological reductionism* is the view that truths about personal identity have as necessary and sufficient conditions statements about the holding of relations of mental continuity and connectedness. Connectedness involves the holding of direct psychological connections, such as the persistence of beliefs and desires, the connection between an intention and the later act in which the intention is carried out, and the connection between an experience and a memory of that experience. Connectedness can come in twice over in the statement of the conditions on personal identity. All psychological reductionists require that, if two person stages are stages of the same person, then psychological continuity, the ancestral of strong or predominant psychological connectedness, holds between them. Some psychological reductionists also require that no two such stages be entirely unconnected psychologically. The dominant view is properly called *wide* psychological reductionism because it has it that mental continuity and connectedness can constitute personal identity even if the holding of these relations is not secured by its normal cause, the persistence of a particular human body or brain. Any causal mechanism that operates so that these psychological relations hold will do. The identity over time of any particular human body or brain plays no strictly indispensable role in the identity of a particular person over time. Any particular human body or brain is just one causal means among others for the holding of the relations of psychological continuity and connectedness which constitute a particular person's survival.

David Lewis, Anthony Quinton, and Sydney Shoemaker are among those who have advanced the *wide psychological view.*[3] The

[2] *Reasons and Persons* (New York: Oxford, 1984), pp. 207/8.

[3] See Lewis, "Survival and Identity," in Amelie O. Rorty, ed., *The Identities of Persons* (Berkeley: California UP, 1976); Quinton, "The Soul," this JOURNAL, LIX, 15 (July 19, 1962): 393–409; Shoemaker, with Richard Swinburne, *Personal Identity* (Oxford: Blackwell, 1984). Shoemaker clearly had some sympathy for the view that bodily continuity is constitutive of personal identity when he wrote *Self-knowledge and Self-identity* (Ithaca, N.Y.: Cornell, 1963), although he allowed that in exceptional cases the bodily criterion of personal identity could be overridden by the memory criterion. He decisively abandons the bodily criterion in "Persons and Their Pasts," *American Philosophical Quarterly*, VII, 4 (October 1970): 269–285. The most notable opponents of the dominant view have been Bernard Williams,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

view is typically defended by appeal to the intuition that one could discover that one had come to inhabit a new body and that others could confirm this by noticing one's familiar personality manifesting itself in that new body. Sometimes this intuition is wrung from cases like John Locke's case of the prince and the cobbler who appear to swap bodies.[4] Sometimes it is wrung from cases of *Star Trek*–style teletransportation in which a body is scanned at one point and is destroyed by the scanning while at another point a cell-by-cell duplicate is made in accord with the information got from the scanning.[5] Sometimes it is wrung from cases of brain-state transfer in which a machine wipes one brain "clean" and another brain is caused to go into states which are identical in all respects to the previous states of the first brain.[6]

The body-change intuition generated by each of these fantastic cases is supposed to support the claim that mental continuity and connectedness, however they are caused, are jointly sufficient for personal identity. In support of the necessity of mental continuity and connectedness it is observed that if the processes in question did not secure these psychological relations we would not be inclined to say that the original person came to inhabit a new body.

The method of cases and the dominant view taken together have a fair claim to constitute the orthodoxy in philosophical discussions of personal identity. Orthodoxies have a tendency to freeze the imagination, so before criticizing this orthodoxy it may be worth while briefly to describe an alternative method and an alternative view.

Suppose that right from the start we give up the reductionist requirement to the effect that statements about personal identity be shown to have the same fact-stating potential as statements about continuity and connectedness. So, in Parfit's terms, we adopt the *further-fact view,* according to which statements about personal identity are made true by facts different from facts about mental and physical continuity and dependence.[7] These latter facts constitute

---

"Are Persons Bodies?," "Bodily Continuity and Personal Identity: A Reply," "Personal Identity and Individuation," "The Self and the Future," all reprinted in *Problems of the Self* (New York: Cambridge, 1973); and David Wiggins, "Personal Identity," ch. 6 of *Sameness and Substance* (New York: Oxford, 1980). My differences with Wiggins are indicated in fn 17 below.

[4] Quinton, *op. cit.* The Locke example occurs in *An Essay concerning Human Understanding,* book II, ch. 17, §15.

[5] Derek Parfit introduces teletransportation on pp. 119/20 of *op. cit.* and discusses it further on pp. 282–287. Parfit, however, claims only that teletransportation is as good as survival. The claim (1) on page 216 *op. cit.*, which seems to imply otherwise, is just misleadingly formulated (personal communication).

[6] Shoemaker discusses the brain-state transfer device in *Personal Identity,* pp. 108–111.

[7] Notice, however, that the account that follows is in no way committed to the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

our evidence for personal identity and are not that in which personal identity consists.

Once we make this first departure from the orthodox position, the method of cases will no longer seem to be the primary method of theorizing about personal identity. We can no longer represent our intuitive reactions to puzzle cases described in the language of continuity and connectedness as mere manifestations of our concept of personal identity. Rather, we will regard those intuitive reactions as judgments based upon the evidence available in the cases. And there will be reason to think that these judgments are more unreliable the more bizarre the cases in question are. For many of the bizarre cases involve the severing of just those contingent connections which in everyday life make continuity and connectedness good evidence for identity.

What then will take the primary place of the method of cases? What are the primary phenomena that a philosophical theory of personal identity should aim to save? Here is a suggestion. There is the humble and ubiquitous practice of reidentifying each other over time. Philosophical skepticism aside, this practice is a reliable and mostly unproblematic source of knowledge about particular claims of personal identity. So the primary question for a philosophical theory of personal identity is: What sort of thing is such that things of that sort can be reliably and unproblematically reidentified over time in *just the way* in which we reliably and unproblematically reidentify ourselves and each other over time? Any theory of personal identity, such as the *bare-locus view* discussed below (sec. II), which would represent our practice of reidentifying each other over time as extremely problematic so far as generating knowledge of the facts of personal identity goes, is automatically ruled out at this stage. And considerable implausibility attaches to any theory that cannot reconstruct as wholly justified the easy and uncomplicated ways in which we reidentify people on the basis of their physical appearance and manner.

---

view that we are "separately existing" (Parfit's phrase) entities distinct from human brains and bodies. Instead I argue that we are human beings and that human beings are constituted by human bodies. Thus the present paper can be read as an extended refutation of Parfit's claim (*op. cit.*, p. 216) that if we believe that personal identity is a further fact we must also believe we are separately existing entities. This claim is crucial to the argument of Part Three of *Reasons and Persons*. Parfit argues directly against the view that we are separately existing entities. But, when it comes to arguing for revisions in our practical attitudes, he takes himself to have established that personal identity is never a further fact. If we are essentially human beings in the sense of the present paper then Parfit's revisionary arguments are broken-backed. Moreover, we can allow Parfit that in some puzzle cases personal identity might well be an indeterminate matter without accepting his revisionary arguments. I elaborate these claims in "Reasons and Reductionism" (forthcoming).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Of course it may be that a number of competing theories survive this first stage. The surviving competitors may then be evaluated in terms of their compatibility with our reactions to those puzzle cases which provide situations in which the competing theories diverge in their pronouncements. But these intuitive reactions are defeasible judgments, which we can defeat by showing that they are overgeneralizations from the ordinary run of cases or are produced by some distorting influence or are outweighed by other judgments that we have reason to respect. So the alternative treatment of intuitive reactions will make more of comparing cases where intuitions conflict and will look for the underlying explanations of these conflicting intuitions. Sometimes the underlying explanations will discredit the intuitions and the conceptions of people which they appear to support, sometimes not. If equally good competitors still survive at this second stage, then we may be faced with genuine indeterminacy which our philosophical theory of personal identity should articulate.

I shall argue that, if we adopt this alternative method which takes a more critical attitude to the method of cases and relegates it to a secondary role, then our investigations will not converge on the dominant view of personal identity. Rather than see ourselves as minds whose *particular* embodiments are contingent, we will see ourselves as human beings: that is, beings which necessarily are normally constituted by human organisms, and whose conditions of survival deviate from those of their constituting organisms only because a human being will continue on if his mind continues on, so that a human being could be reduced to the condition of a mere brain so long as that brain continues the human being's mental life. Although having all but its brain destroyed is too much for a human organism to survive, a human being might plausibly be held to survive this. But as I (somewhat stipulatively) shall use the term 'human being', no human being could survive teletransportation or like cases of *complete* body transfer.[8]

[8] Although, for the purposes of this paper the possibility is set aside to avoid complications, I do not rule out a kind of relativism about personal identity which allows that acculturated human animals blamelessly could have had or could come to have a conception of personal identity according to which "they" could survive teletransportation. (Whether the scare quotes can be taken off ' "they" ' is of course a moot point.) The main claim of the present paper is that there is no reason to think that our actual conception of personal identity is such as to allow us to survive teletransportation. The main impact of relativism about personal identity and the possibility of refiguring our conception of ourselves will not be that we are not of the kind *human being*, but that this kind may not strictly be a *substance* kind, i.e., may not fix in an absolutely invariant way what kind of changes we could survive.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

By way of beginning to displace the orthodox position—the method of cases and the dominant view—I shall present a well-known conundrum about personal identity which, on the face of it, seems to threaten both parts of the orthodox position. A relatively orthodox reaction to the conundrum will be found wanting. In extracting the right morals from the conundrum I shall employ the alternative methodology outlined above. It will then become evident that the dominant view depends for its appeal upon the exploitation of certain systematically distorting influences on our intuitions about puzzle cases.

I. THE CONUNDRUM

The conundrum that threatens both the method of cases and the dominant view derives from Bernard Williams,[9] though I will present it in my own way.

In the fantastic tale that follows we are to suppose that there is a machine which produces the sort of psychological effects that would be produced by transposing patients' brains. When the machine is connected up to two patients, *A* and *B*, and turned on, the machine records all features of their brains relevant to the determination of dispositional and occurrent mentality. The machine alters the *A*-brain so that it comes to have associated with it what appears to be the continuation of *B*'s dispositional and occurrent mentality and so that the person with the *A*-body emerges from the machine with what appear to be *B*'s course of experience, memories, character traits, projects, etc. Mutatis mutandis for the *B*-brain. The person with the *B*-body emerges from the machine with what appear to be *A*'s course of experience, memories, character traits, projects, etc.

Suppose that *A* is faced with such a procedure. First his psychology is realized in his brain and body. Then the machine is turned on and his psychology appears to be realized in the *B*-body and brain. Mutatis mutandis for *B*. Before the machine operates it appears to *A* that the machine's operation would cause him to swap bodies with *B*. So if before the machine's operation *A* were to choose which of the *A*-body and the *B*-body is to undergo pain after the machine's operation and choose only from the self-interested point of view, he would choose the *A*-body. When the case is presented this way many would agree with *A*. After all, if the brains of *A* and *B* were surgically

---

However, on the assumption that the radical changes in our thought and attitudes required for refiguring our conception of ourselves will not take place, we may treat the kind *human being* as if it were our substance kind and so as if it determined our essence. On these matters see my "Relativism and the Self," forthcoming.

[9] "The Self and the Future," *Philosophical Review*, LXXIX, 2 (April 1970): 161–180.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

transposed and the resultant composites survived, then, so many of us believe, *A* and *B* would have swapped bodies. The machine seems to produce the same effects without having to transpose brains.

But now suppose that *A* has to choose whether to undergo a very painful surgical procedure which has a small chance of curing him of some illness. There is no effective anesthetic or analgesic that will block the pain. But *A*'s doctor tells him that the latest medical technology can be used to "get around" the pain. The doctor explains that he is able to record all the features of *A*'s brain relevant to the determination of dispositional and occurrent mentality, and reorganize those features so that dispositional and occurrent mental features very different from *A*'s will be realized by *A*'s brain. Then and only then will *A*'s body be subjected to the painful surgical intervention. Only after the pain has ceased will the machine reorganize *A*'s brain so that it realizes dispositional and occurrent mentality continuous with *A*'s mentality as it was before the machine first acted on his brain.

Presented with such a prospect and asked to regard the future pain as having no significance from the self-interested point of view, *A* might reasonably retort that he is being asked to undergo a double assault. First his brain is to be fiddled with in a fairly drastic way so as to produce radical psychological discontinuity, and then he is to be caused to feel severe pain. And this reaction is in accord with the intuition most of us would have about the case. The future pain is *A*'s pain despite the intervening psychological discontinuity. As Williams notes, the concept of a person's future pain is not such that some degree of psychological continuity is required in order to make some future pain his.

So we have two intuitions, the intuition that *A* would swap bodies and so avoid any pain then undergone by the *A*-body, and the intuition that *A* would suffer pain in the *A*-body despite radical psychological discontinuity. The second intuition counts against the dominant view, which takes psychological continuity to be necessary for personal identity. And, on reflection, a further difficulty emerges. We have two directly conflicting intuitions about what could be taken to be two presentations of the same case. For we may suppose that the medical technology referred to by the doctor is just the machine that appears to transpose minds. What the doctor described was just one side of the body-swapping case followed by another body swap to reinstate *A* in the *A*-body and *B* in the *B*-body. The doctor refrained from describing *B*'s experiences to *A*; but just how can this be relevant to the determination of our intuitions in the second presentation of the case?

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Taken at face value, the intuition associated with the second presentation: that *A* would suffer pain in the *A*-body despite intervening psychological discontinuity of the most radical sort, threatens the dominant view. And on the face of it the two intuitions taken together threaten the method of cases. For how can intuition be reliable if we can be got to react so differently to the very same case?

Robert Nozick has recently offered a diagnosis which, if successful, would save the method of cases and only slightly modify the dominant view. He suggests that our intuitions about personal identity conform to the *closest-continuer schema* according to which the closest of the sufficiently close continuers of a given individual is that individual.[10] This is properly called a "schema" because just which weightings of just which relations are relevant in determining closeness of continuation will vary depending on what kind of individual is in question. The intuition associated with the second presentation can be taken as showing that, when it comes to persons, bodily continuity can make for sufficient continuity in the absence of psychological continuity. This explains our tracing *A* so that he remains in the same body throughout. The intuition associated with the first presentation can be taken as showing that when bodily continuity and psychological continuity diverge we give more weight to psychological continuity. This explains why, when told of the adventures of *B*'s psychology as well as *A*'s, we have the intuition that *A* and *B* swap bodies. Our intuitive reactions to the two presentations are consistent because the first presentation simply describes a better continuer of *A* than any continuer of *A* described in the second presentation. That is why mentioning *B* and what happens to his brain is relevant. And the dominant view requires only modification. Psychological continuity can be on its own sufficient for personal identity and, in the absence of bodily continuity, necessary for personal identity.

This would be a satisfying orthodox response to the conundrum if it could be shown that our intuitions about personal identity conform in general to the closest-continuer schema interpreted in terms of the modified dominant view. But they do not.

Suppose the brain-state transfer machine is developed further so that it can seem to continue the mentality of a given patient *A* in each of the bodies of two patients *B* and *C*. Typically, the original *A*-body dies as a result of the machine's operation, and, after the machine operates, the person associated with the *B*-body and the person associated with the *C*-body are each sufficiently close psychological

[10] *Philosophical Explanations* (Cambridge, Mass.: Harvard, 1981), pp. 29–70.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

continuers of *A*. Suppose also, however, that typically the person associated with the *C*-body is a considerably better psychological continuer of *A* than the person associated with the *B*-body. Finally, suppose that this extra continuity is more than enough to compensate for the extra ten-minute delay typically involved in "reading" *A*'s psychology into the *C*-body.

Now imagine that the machine has been operating and *A*'s psychology has been (or seems to have been) "read into" the *B*-body. The person associated with the *B*-body gets up and walks around and thinks to himself, "I am *A*. I did not just come into existence." He sees the machine beginning to "read" *A*'s psychology into the *C*-body. He knows that if the process is allowed to continue the result will be that the person associated with the *C*-body will be a better continuer of *A* than he is. So he turns off the machine. According to the closest-continuer theory the person associated with the *B*-body has made it true that he is *A* and has existed before the operation of the machine as *A*. But, as against this, surely our intuition is that the *B*-body person's thought, "I am *A*. I did not just come into existence," is made true or false by what has happened up to and including the time at which that thought occurs. Surely no subsequent act by the *B*-body person can make this thought true or false. The closest-continuer schema interpreted in terms of the modified dominant view violates the intuition that the relevant facts are settled before the *B*-body person decides whether or not to tamper with the machine. To the extent that we have this and like intuitions, we are not to be construed as responding to cases as if we were adherents of the closest-continuer schema, at least as applied to people. To that extent Nozick's reaction to the conundrum, a conundrum which derives from our present responses to two presentations of the same case, is inadequate, *even if it were the case that philosophical theorists should ultimately adopt the closest-continuer schema in some form in the light of all the evidence about personal identity.*

There is another consideration that suggests that Nozick's diagnosis of the conundrum is inadequate. That diagnosis crucially depends upon the second presentation of Williams's case being one-sided, in the sense that the only sufficiently good continuer of *A* presented is the *A*-body person. The closest-continuer schema applied in accord with the modified dominant view implies that if the second presentation had included reference to another body who appears to inherit *A*'s psychology then we would simply have the first presentation and the familiar body-swap intuition. But in fact we easily understand the *stipulation* that *A undergoes* or survives throughout the "reading into" his brain of *B*'s psychology and then

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

suffers severe pain while *B undergoes* the "reading into" his brain of *A*'s psychology. An adequate response to the conundrum has to explain why we can make sense of this *stipulation* even though we respond as we do in the first presentation of Williams's case.

What then can we say about the conundrum? Here is a minimal response. (Competing elaborations of the response will be considered below.) It seems that our reactions to the two presentations indicate that we think of a person as a *locus* of mental life; that is to say, something which typically exhibits psychological continuity and can be traced in terms of it, but something which need not always exhibit psychological continuity, something which can undergo psychological discontinuity of the most radical sort. The first presentation has us trace *A* in terms of psychological continuity, the second presentation encourages us to suppose that *A* undergoes psychological discontinuity. Our reactions are comprehensible if we think of ourselves as loci of mental lives, where this characterization does not in itself require that we think of ourselves as things of such a sort that psychological continuity, however it is secured, is sufficient for our survival. All that needs to be allowed is that it is reasonable to trace a locus of mental life in terms of psychological continuity. It is left open whether a reasonable practice might be leading us wrong in the puzzle case at hand.

Given this way of understanding our reactions to the conundrum, we can take our reaction to the second presentation as prima facie evidence against the wide psychological view (since such a view would make no sense of the stipulation that *A undergoes* radical psychological discontinuity) without automatically taking our reaction to the first presentation as showing that psychological continuity, however it is secured, is sufficient for our survival. At least, this is so if we are able satisfactorily to answer the question: Just what *kind* of locus of mental life are we? in a way that does not imply that psychological continuity, however it is secured, is sufficient for our survival. This is the question that will have to be answered by any theorist of personal identity whether or not he aims for a reduction of the persistence conditions of people in terms of conditions on continuity.

II. WHAT KIND OF LOCUS OF MENTAL LIFE?

The dominant view, or wide psychological reductionism, amounts to a fairly specific characterization of the kind of thing we are, since it has psychological continuity and connectedness figure as necessary conditions on personal identity. But, given its reliance on the method of cases, the dominant view has no clear right to this necessary condition and consequent degree of specificity. For if the method of cases is taken at face value we get the result that there are no specific

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

necessary conditions on the survival of a person. That is, Williams's conundrum and like cases could be taken to show that we are what I will call *bare loci* of mental life, that is, possessors of mental life whose survival requires no amount of either bodily or mental continuity.[11] If the method of cases is taken as *the* method of theorizing about personal identity such a view can be made to appear quite compelling. It is only when we adopt the alternative method of theorizing about personal identity described above that the bare-locus view can be *shown* to be unacceptable.

The first presentation of Williams's case seems to show that bodily continuity is not necessary for survival, and the second presentation seems to show that psychological continuity is not necessary for survival. We may bolster up these appearances by pointing out that we can imagine many sorts of cases that seem to involve one's ceasing to be associated with a particular human body and human personality. These cases are particularly compelling when imagined "from the inside." So I am to imagine myself undergoing a radical change in my form, e.g., a change like that undergone by Franz Kafka's beetle-man,[12] and perhaps concurrently a wild change in my psychology. There seems to be nothing internally incoherent about such imaginings. None of them exhibit the sort of conceptual deadlock associated with the attempt to imagine a married bachelor or a round square (if anything could be called such an attempt). Nor do such imaginings appear necessarily to embody some religious or secular conception of people which the imaginer has picked up. Rather it seems that such imaginings are not ruled out by and so are consistent with our concept of a person. So, given the ideology behind the method of cases, it is difficult to dismiss such imaginings as idle. They will appear to chart the boundaries of our concept of a person and thus indicate the correct theory of personal identity.

But such imaginings are idle. And if this can be shown, then it is plausible to take our pure *concept* of a person, understood as what-

[11] Geoffrey Maddel argues, in *The Identity of the Self* (Edinburgh: University Press, 1981), that our survival requires no amount of either mental or bodily continuity, taking himself to be arguing in the tradition of Butler, Reid, and McTaggart. See especially pp. 117–140. I hesitate, however, to attribute the bare-locus view to Maddel, since he insists that people are not objective entities at all, but rather thoroughly "subjective" in the sense of Thomas Nagel, "Subjective and Objective," in *Mortal Questions* (New York: Cambridge, 1979). It may be that this idea of ourselves as subjective is a seductive Idea of Reason in something like the Kantian sense, an idea which derives its apparent plausibility from the fact that the only conception of ourselves that reflective "I"-thought appears to underwrite is the unspecific conception of ourselves as some or other sort of locus of reflective mental life. See what follows and fn 16.

[12] *The Metamorphosis* (*Die Verwandlung*), Willa and Edwin Muir, trans. (New York: Schocken Books, 1968).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ever it is that our reactions to these puzzle cases manifest, as being too unspecific to be of much interest, and as unspecific in a way that will typically lead us wrong in many puzzle cases.

Suppose, for example, that our pure concept of a person is *just* the concept of an unspecified kind of locus of (perhaps reflective) mental life, where this is understood, as before, to allow that a locus of mental life can survive radical psychological discontinuity. This concept is unspecific in the sense that it embodies no determinate necessary conditions on the survival of a person. There are no determinate necessary conditions on any tracing of a locus of mental life through any imagined vicissitude. Of course, in the *presence* of imagined bodily or mental continuities, we typically trace in terms of them, but this is just the reflection in our imaginative conceits of our ordinary practice of tracing people. That practice employs such continuities as evidence for the survival of people, and some of our reactions to imagined puzzle cases can be taken as harmless extensions of this practice. But, in the *absence* of imagined bodily or mental continuity, our unspecific conception of people is unconstrained by any evidence for persistence, and we get just the intuitive responses that seem to indicate that we are beings of the kind *bare locus of mental life,* where this characterization is now understood as including all that is necessary and sufficient for our persistence through time, so that we could undergo any sort of physical or mental metamorphosis.

How are we to show that the imaginings that detach me from the human being I appear to be are idle, without begging the question against the bare-locus view? The alternative method outlined above lays down as a first constraint on any theory of personal identity that it enable us to reconstruct our everyday practice of reidentifying people as an unproblematic source of knowledge about personal identity over time. This is a constraint which the bare-locus view flouts.

To see this, consider one of the wilder imaginative conceits which is identity-preserving on the view that I am a bare locus. I can imagine something that might be called my experiencing my body turning to stone. A queer numbness creeps up from my feet. It is as if my body is getting heavier and heavier and less and less tractable to my will. My senses give out and in the end I am in the grip of panic at what seems like complete paralysis and sensory loss.

If we suppose that I can survive my body turning to stone we must then face the question of which relations tie me to my body. The supposition that I could survive my body's petrification implies that the relations that tie me to my body are contingent. It cannot be that I am identical with or necessarily constituted by my living body. It

CC-0.In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

seems then that the relevant relation can at most involve some actual pattern of causal interaction or some particular channel of causal influence (which may or may not be exploited at any particular time) by which I, a bare locus, get information about and direct changes in some of the states of that body.

The supposition that it is always some actual pattern of causal interaction which ties the bare locus that I am to a particular body can be set aside. If I can survive my body turning to stone because my essence is to be a bare locus of mental life, then surely I could survive my body turning to stone when that body is in deep (dreamless) sleep or rendered unconscious by drugs that work on the brain. For there was nothing in the idea of a locus of mental life as we introduced it which required that such a locus never undergo an interruption in its occurrent mental life. Otherwise the bare-locus view would be at odds with what appear to be the facts about our surviving periods of unconsciousness. Suppose then that, while it was unconscious, the living thing that I call "my body" turned to stone. The contingent relation which held between the bare locus that I am supposed to be and my body before it turned to stone and which ceased to hold when my body turned to stone was not some actual pattern of causal interaction by which that bare locus got information about and directed changes in some of the states of my body. None of that was going on when I was unconscious.

So we are driven back to the suggestion that, according to the bare-locus view, what makes a particular human body my body is a channel of causal influence which holds between the bare locus that I am and the body that I call "my body." But now the epistemological difficulties associated with the bare-locus view come to the fore.

Ordinary judgments that a particular person is in a state of unconsciousness are based on the observed continuity of human bodily life. Such judgments typically constitute knowledge that the particular person still exists though unconscious. If one condition on knowing that p is having evidence that converges on p as opposed to the relevant alternatives to p then, in cases of unconsciousness, continuity of human bodily life must favor the proposition that a single person's survival is manifested by that bodily life. But then, in typical cases of unconsciousness or sleep, there cannot be a relevant alternative to the effect that people understood as bare loci have been replacing each other behind the stage of one human body's life. For that is an alternative that we are not in a position to rule out when faced with a sleeping or unconscious human body. If such an alternative were relevant we would have little chance of knowing who was in bed with us at night.

For suppose that one came to accept the bare-locus view, com-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

plete with the account of channels of causal influence tying bare loci to particular human bodies. Then the ordinary judgment that a particular person is still in bed even though in deep sleep will have to be construed as the judgment that the same bare locus of mental life is associated with the living body in question via a channel of causal influence which is not then being exploited. But such a judgment brings in its train a host of alternative hypotheses, e.g., that there are two or three or four or many more such loci, perhaps each with its appropriate channel, replacing each other behind the stage of the continuous bodily life that we observe. Given the *metaphysical* picture of a bare locus being tied to a body by a channel of possible causal influence, these *become* relevant alternatives. Let no one suggest that those ignorant of the ways of bare loci can rule out these alternatives by means of an inference to the numerically *simplest explanation* of observed bodily continuity. None of these hypotheses *explain* continuity of a bodily life devoid of mental accompaniments. They are all merely compatible with such continuity. And indeed one should want to know why the simplest hypothesis is not that in deep sleep the bare locus usually tied to a given body has faded away or has simply detached itself from that body to wander where it will, as in the spiritualist's fantasies of astral travel. The relative simplicity of such alternative hypotheses could be reasonably evaluated only in the presence of some account of the typical or natural ways of bare loci.

Suppose there were some empirical theory that one could offer about bare loci, perhaps confirmed by introspection and observation of others or even by revelation, and—the best case—such a theory implied that bare loci did not typically migrate or permute during unconsciousness or sleep. This is roughly what any substance dualist, the bare-locus theorist included, should take himself to be offering, namely, an empirical theory that fits the facts of human experience.[13] By appeal to the theory one might have a way of ruling out the relevant alternatives and so a way of coming to know that a single person is associated with a sleeping or unconscious body over a period of time.

[13] That is, I reject attempts to rule out every form of substance dualism a priori in the fashion of P. F. Strawson, *Individuals* (London: Methuen; New York: Doubleday, 1964), pp. 90–103, or, more recently, Jay Rosenberg, *Thinking Clearly about Death* (Englewood Cliffs, N.J.: Prentice Hall, 1984). On this see Shoemaker, "Two Kinds of Dualism," in his collected papers, *Identity, Cause and Mind* (New York: Cambridge, 1984). On Donald Davidson's attempt to rule out a priori a dualism of mental and physical events, see my "Why Having a Mind Matters," in Ernest LePore and Brian McLaughlin, eds., *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (New York: Blackwell, 1985).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The crucial point is that our ordinary claims to know that our friends and familiars were continuously where their bodies were when they were unconscious or in deep sleep rests on nothing like the employment of any such theory to rule out the possibility that any number of bare loci came to be associated with their bodies during such periods. These are alternatives we just do not consider, let alone rule out. We would boggle if asked to take them seriously. Yet they would be relevant alternatives if the bare-locus view were the correct view of people. So if the bare-locus view is correct then our ordinary practice of reidentifying people during periods of unconsciousness and deep sleep does not generate knowledge. But, philosophical skepticism aside, we have every reason to think that it does.

One cannot save the bare-locus view by holding that, although it is the true view, in ordinary life we are spared the obligation to rule out the alternatives it makes relevant by our ignorance of the fact that it is the true view. For this sort of ignorance does not make it easier to know. Take just one familiar example. Suppose that my interlocutor and I both see a woman who looks just like Mary steal a book from the library. It is in fact Mary, and in fact I believe this, but I am not in a position to know this because I know that Mary has an identical twin who dresses just like her, so there is a relevant alternative that my observational evidence does not rule out. My interlocutor knows nothing of Mary's twin and claims to know that Mary stole the book on the strength of what we both saw. I take it that our intuition is that, even if it was Mary we both saw, his claim to know is mistaken, for he cannot know and I not know just because he is ignorant of crucially relevant facts.[14]

If this is right, the bare-locus view flouts the first constraint of the alternative method. Given that view, we cannot reconstruct a part of our ordinary practice of reidentifying each other as generating knowledge about personal identity. If the bare-locus view fails, a number of views fail with it. For similar arguments would count against any substance dualist, such as Richard Swinburne, who identifies us with mental substances, even if that substance dualist allows that we have some mental features as permanent or essential properties.[15] On any such view similar problems will arise with unconsciousness and deep sleep. We do not trace a locus of mental life with a purely mental nature or essence when we trace people through such states.

[14] For similar cases and a discussion of relevant alternatives, see Alvin Goldman, "Discrimination and Perceptual Knowledge," this JOURNAL, LXXIII, 20 (Nov. 18, 1976): 771–791, and Marshall Swain, "Reasons, Causes, and Knowledge," *ibid.*, LXXV, 5 (May 1978): 229–249.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

[15] See "Personal Identity," *Proceedings of the Aristotelian Society*, LXXIV (1973/4): 231–247, and Shoemaker and Swinburne, *op. cit.*

Digitized by Arya Samaj Foundation Chennai and eGangotri

The bare-locus view is straightforwardly supported by the method of cases. To the extent that we find that view absurd in its consequences, we should have grave doubts about relying simply upon the method of cases. For if we are not bare loci of mental life or activity, then some of our intuitions in puzzle cases simply represent our tracing what is no more than a reification of our unspecific concept of ourselves as *some or other sort* of locus of mental life.[16] But what is the correct specific conception of ourselves, and how could we argue for it without running foul of intuitions generated by a misleading reification of this unspecific concept of ourselves?

III. HUMAN ORGANISMS AND HUMAN BEINGS

Suppose that, instead of engaging in the self-defeating search for truths about ourselves that are both purely conceptual and yet specific enough to be interesting, we start with the specific conception of ourselves as evolved animals of a particular sort, the conception that locates us most easily within the naturalistic framework taken for granted by scientifically validated common sense. We will then try to articulate which aspects of our thought about ourselves could lead us to displace or modify this conception. The point of departure is then the position that we are essentially organisms of a particular animal species, namely *homo sapiens,* so that the locus of mental life that we reidentify when we reidentify a person over time is just an instance of a biological kind, a kind whose members typically exhibit a complex mental life.

That we must depart from this position is, I think, the upshot of the case first introduced into philosophical discussion by Sydney Shoemaker:

> It is now possible to transplant certain organs . . . it is at least conceivable . . . that a human body could continue to function normally if its brain were replaced by one taken from another human body . . . Two men, a Mr. Brown and a Mr. Robinson, had been operated on for brain tumors, and brain extractions had been performed on both of them. At the end of the operations, however, the assistant inadvertently put Brown's brain in Robinson's head, and Robinson's brain in Brown's head. One of these men immediately dies, but the other, the one with Robinson's head and Brown's brain, eventually regains consciousness. Let us call the latter "Brownson" . . . When asked his name he automatically replies "Brown." He recognizes Brown's wife and family . . . ,

[16] Compare Kant's diagnosis in the Paralogisms (B427) of the Cartesian mistake: "I think myself on behalf of a possible experience, at the same time abstracting from all actual experience, and I conclude therefrom that I can be conscious of my existence even apart from experience and its empirical conditions. In so doing I am confusing a possible abstraction from my empirically determined existence with the supposed consciousness of a possible separate existence of my thinking self, and I thus come to believe that I have knowledge that what is substantial in me is the transcendental subject."

CC-0. In Public Domain. Panini Kanya Maha Vidyalaya Collection.

Digitized by Arya Samaj Foundation Chennai and eGangotri

and is able to describe in detail events in Brown's life . . . of Robinson's past life he evidences no knowledge at all (*Self-knowledge and Self-identity,* 23/4).

The predominant reaction to this case is to suppose that Brownson is Brown and that during the brain-switching operation Brown survives as his denuded or debodied brain. This intuition is robust in the sense that it remains even if we modify the case by supposing that Brown's debrained body—call it "Brownless"—is provided with enough in the way of transplanted brain-stem tissue to keep it alive indefinitely. Although the transplanted brain-stem tissue does not subserve consciousness, Brownless is alive. Blood is coursing through Brownless, compounds in that blood are being metabolized, breath is being drawn into the lungs of Brownless, wastes are being expelled, and so on. Brownless is a living though badly mutilated human organism. And it seems bizarre to suppose that the organism Brownless came into being only after Brown's brain was taken out of Brown's body. Instead it seems that Brownless is the same organism as Brown. If we trace Brown's original organic life we find Brownless realizing most of this life after Brown's brain is transplanted into Robinson's debrained body. We can allow a sense in which Brown's mental life may seem to continue on in Brownson. But, given the predominance of basic life functions that continue on in Brownless, we cannot count Brownson as the same human organism as Brown *and* think of human organisms as instances of purely *biological* kinds; i.e., individuals whose persistence through time is constituted by continuous biological life, so that the relative importance of life functions (including mental functioning) in tracing such an individual is determined by the relative contribution that those functions make to maintaining continuous life. For in this respect metabolism is more important than mentation. So if 'human organism' is taken to pick out a purely biological kind, Brownless and not Brownson is the same human organism as Brown. Hence if Brown survives as Brownson, neither Brown nor his kind-mates (those essentially the same as he) are essentially human organisms.[17]

[17] Hence Shoemaker's case raises problems for David Wiggins's account of human persons in *Sameness and Substance, op. cit.* Wiggins suggests, "a person is any animal that is such by its kind to have the biological capacities to enjoy fully the psychological capacities enumerated," and extracts this consequence: "There would be no one real essence of person as such, but every person could still have the real essence of a certain kind of animal. Indirectly this would be the real essence in virtue of which he was a person" (172). If Brown is not the same human organism as Brownson then he is not the same animal as Brownson. So if Brown survives as Brownson, then the animal kind *human animal* does not capture Brown's essence. Similar remarks apply to Peter van Inwagen's claim that persons are organisms, a centerpiece of his *Material Beings* (Ithaca, N.Y.: Cornell, 1987).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Usually, what I've called the "predominant" reaction to Shoemaker's case, *viz.*, the fact that most judge that Brownson is Brown, is taken as establishing that Brownson is Brown. But this is just another example of uncritical reliance on the method of cases. Indeed, the predominant reaction can be explained in terms of the distorting influences described below as the psychological- and social-continuer effects. *We* need a principled reason, consonant with the alternative method, for taking seriously the judgment that Brownson is Brown and thereby departing from the straightforward naturalistic conception of ourselves as essentially human organisms.

The alternative method requires that we represent ourselves as things of a kind such that the easy and uncomplicated ways in which we ordinarily trace people are well adapted to tracing things of that kind. Now, in tracing oneself backwards in time, one typically relies upon experiential memory. The deliverances of experiential memory are propositions involving claims about personal identity, *viz.*, that the one who is now having the memory experience is the one whose experience is being remembered. That is, experiential memory does not leave us at a halfway house short of a full-blooded belief about personal identity. It does not leave us with some inference to make to arrive at the belief that it was none other than the one who is now having the memory who had the experience being presently remembered. Moreover, it seems that, even if one were in the Cartesian predicament of having suspended belief in anything external to one's own mind, experiential memory could still deliver knowledge about personal identity.

This raises a difficulty. Once it is admitted that among the necessary conditions on personal identity are certain bodily conditions having to do with the survival of an organism or crucial parts of it, the question arises how experiential memory, i.e., something whose internal phenomenology makes it seem like a faculty suited to picking up only mental connections between earlier and later mental states, could deliver any more than conclusions of the form: the same *mind* that had the remembered experience is the mind that is now remembering. For how could such a faculty directly pick up the fact that some bodily condition on personal identity is satisfied? And if it could not do this, by what right are its deliverances properly taken to involve claims about personal identity as opposed to claims about mere mental identity?

The way out of this difficulty for the reconstructive program that is the alternative method is to realize that, if anything deserves the name of a conceptual truth about the relation between persons and minds, it is the claim that a person cannot be outlived by (what once was) his own mind. It is not a temporary feature of my mind that it is

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

my mind. Nor could it be. No situation could deserve a description to the effect that the very mind that is my mind has been or will come to be the mind of someone else. Talk of a particular mind is just talk about a particular person's mental functioning. So, in delivering the conviction (and typically the knowledge) that the mind that had the remembered experience is the same mind that is now remembering, experiential memory is *eo ipso* delivering the conviction (and typically the knowledge) that the person who had the remembered experience is the same person as the person who is now remembering.

If this is so, then, whenever we have reason to say that a single mind has continued on, we have reason to say that a single person has continued on. But if we now adopt a properly naturalistic view of our mental functioning, i.e., see our mental functioning as the characteristic functioning of our brains, then it will be difficult, albeit not impossible, to resist the idea that one's mind would continue on if only one's brain were kept alive and functioning. Given the conceptual connection between people and minds, this amounts to the conclusion that one would go where one's brain goes and that one could survive as a mere brain. And now we have not just a brute intuition but an argument for the predominant reaction to Shoemaker's case from a principle that must hold if experiential memory is to be reconstructed as a faculty that could deliver knowledge about personal (as opposed to merely mental) identity. Applying the alternative method and attempting to reconstruct our reidentifications of others, things go most smoothly if we understand ourselves as tracing human organisms. But once we try to reconstruct our reidentifications of ourselves via memory, we see that it is human beings that we trace, i.e., beings that could outlive the human organisms they are invariably constituted by if their minds were to continue on.

Focusing on intuitions about brain-transfer cases and, in particular, the intuition that during such processes one survives as a mere brain has led some philosophers to take seriously a view floated for discussion by Thomas Nagel and J. L. Mackie, the view that we are essentially human brains, so that, for example, 'I' in my mouth picks out a pink, spongy thing inside my skull.[18] This extraordinary view is of the same ilk as William James's notion that his true self was located somewhere near the back of his throat and G. E. Moore's conviction that he was closer to his hands than to his feet. Rather than argue directly against the view that I am my brain and its strange implications (e.g., that when I report my weight as 150 pounds I am like a

[18] Nagel, "Are You Your Brain?," paper delivered to Princeton Philosophy Colloquium and APA Pacific Division, 1984. Mackie, "The Transcendental 'I' ", in Zak van Straaten, ed., *Philosophical Subjects: Essays Presented to P. F. Strawson* (New York: Oxford, 1980).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

driver of a heavy truck who says to a bridge attendant "I weigh 3 tons") it is better to show that what is correct in the motivation for the view can be accommodated without needless paradox or wriggling out of paradox. The plausible observation behind the Nagel/Mackie view is that the survival of one's brain can be sufficient for one's survival and may well be necessary for it. But this does not show that we are of the kind *human brain,* i.e., that in our normal unmutilated condition we are pinkish-grey, spongy organs awash in cerebrospinal fluid. For the conditions under which one of us might be held to survive as a mere brain are just the conditions where it would make sense to talk of a *radically mutilated* human being, one reduced to the condition of a mere brain. The concept of mutilation must be respected here; we cannot determine the characteristic extent and form of a human being by determining how much mutilation it *can* undergo. The characteristic extent and form of a human being is that of an unmutilated human being, and that is the extent and form of an unmutilated human organism. This obvious claim is in no way threatened by the observation that each human being has a proper part, his brain, which is such that the survival of that part may constitute his survival in a radically mutilated condition and which is such that he would not survive the destruction of that part.

Hence allowing that in cases like Shoemaker's one would go where one's brain goes does not imply that one is of the kind *human brain,* but only that one is of a kind such that its members survive if their mental life continues on as a result of the survival of their organ of mentation. The kind *human being* is such a kind if the tracing of the life of a human being gives primary importance to mental functioning among the various life functions exhibited by human beings. In this sense 'human being' names a partly psychological kind, whereas 'human organism', as we were understanding it, names a purely biological kind, a kind such that mental functioning is given no special persistence-guaranteeing status among the various life functions exhibited by the instances of that kind. Nevertheless, these kinds are intimately related. So far in every actual case a human being is constituted by a particular human organism, and so the survival of the organism is for all practical purposes a necessary condition of the survival of the human being. There are possible cases, such as Shoemaker's, in which a human being can come to be constituted by a mere brain and then by another organism which comes to have that brain as its organ of mentation. But it is crucial to the tracing of a human being that there be something that is the continued functioning of that human being's brain, and so a human being cannot survive teletransportation and the like. The fact that teletransportation secures certain causal dependencies between earlier and later

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

mental states should not seduce us into thinking that these are states of the same mind. A human mind is neither an independently traceable substance nor some bundle theorist's ersatz for such a substance. A human mind is just a mode of functioning of a natural unit (e.g., a human organism or a human brain) whose conditions of persistence are statable in nonmental terms.[19] This is the sense in which talk of a mind is overly reified talk of an aspect of some minded thing.

IV. THE CONUNDRUM AGAIN

We have some final work to do on Williams's conundrum. If we are essentially human beings, we can undergo radical psychological discontinuity as a result of having our brains tampered with, but we cannot switch bodies without brain transplantation. So the view that we are essentially human beings is incompatible with the intuition generated by the first presentation of Williams's case. What is going on here?

We have found good reason to suspect intuitions that show that one or another form of continuity is not necessary for survival; for such intuitions may simply be generated by an unspecific concept of ourselves as some or other kind of locus of mental life. That having been said, it should be noted that, in the first presentation of Williams's case, we trace *A* in terms of psychological continuity even though we have the option of tracing *A* in terms of bodily continuity. This is not to be explained simply in terms of the employment of an unspecific concept of ourselves as loci of mental life. Why do we favor psychological continuity over bodily continuity? Is it that we are implicitly committed to the wide psychological view?

I think not. Our tendency to trace people in terms of psychological continuity in those puzzle cases in which such continuity comes apart from bodily continuity can be accounted for more satisfactorily as an understandable overgeneralization from the ordinary run of things. Let us say that *X* is an excellent continuer of *Y* if *X*'s occurrent mental life evolves out of *Y*'s occurrent mental life and *X*'s dispositional psychology (character traits, quasi memories, beliefs, and desires) evolves out of and is very similar to *Y*'s dispositional psychology, so that *X* will experience and be disposed to do what *Y* would have experienced and have been disposed to do. Now in everyday life the normal concomitant of death is the loss of any very good contin-

[19] The defense of this claim and the task of giving an acceptable account of constitution, persistence, and kinds I leave for another occasion. On the matter of kinds all that is needed for present purposes is that all actual and possible members of a kind share the same general conditions of persistence through time and the same possible types of constitution over time.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

uer of oneself. When one dies there is no one to take one's place in the world, to draw on one's memories and experiences, to act on one's intentions, to do the things one would have done in just the way one would have done them. And the normal concomitant of survival is having over the short term a unique and excellent continuer, namely oneself.

Now if we are essentially human beings, it will be possible to imagine cases in which a person ceases to be and yet nonetheless has a unique excellent continuer and also cases in which a person continues to exist while another becomes his unique, excellent continuer. Parfit's teletransportation case will be a case of the first sort, whereas Williams's first presentation depicts a case of the second sort. If such cases are described simply in terms of continuities we will be liable to be misled by the normal psychological concomitants of survival and so trace individuals in terms of psychological continuity. By means of an understandable overgeneralization from the ordinary run of cases we will be led to trace individuals in accord with the wide psychological criterion. We could call this overgeneralization *the psychological-continuer effect.*

This can only be part of the explanation. The question remains why, in reacting to the first presentation of Williams's case, we do not instead generalize from another typical "concomitant" of survival, namely, the fact that in everyday life one has the same body throughout any process that one survives. Part of the answer is that people *can* be got to react this way to the first presentation of Williams's case if it is described immediately after cases that highlight the importance of bodily continuity. But when I described the first presentation I followed Williams and deliberately assimilated it to a brain-transplanting case, describing it as a case that produces the same psychological effects as brain transplanting without the surgical messiness of an actual brain swap. Given the concurrent presentation of a brain-swap case, the sort of bodily continuity that is typical of survival in everyday life is made less salient, and the psychological-continuer effect tends to operate. But in my experience it is difficult to get the uninitiated to come up with the body-swap intuition in response to the first presentation of Williams's case when the first presentation comes immediately after cases that emphasize the importance of bodily continuity. Such inconstancy in our intuitive reactions itself suggests that our ordinary capacity to make correct judgments about personal identity is not well engaged by such bizarre cases. We are bewildered, and in our bewilderment we opt for one or another partial extension of our ordinary practice of reidentification.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

It must be admitted, however, that there is a tendency among some people to respond to the first presentation of Williams's case with the body-swap intuition, even when the first presentation is delivered on its own and without any suggestion about which other cases it should be assimilated to. This is especially so when, in the case presented, *B*, the recipient of what seems like *A*'s psychology, is like *A* in most bodily respects. The question remains why for some, in such a case, the psychological-continuer effect still predominates over the bodily-continuer effect.

Two factors may play a role. First, our practical interests in the relation of personal identity: in everyday life the persistence of a person typically guarantees that over the short term there will continue to be some occupant of the particular complex of detailed and manifold social roles (father, lover, friend, leader, supporter, colleague, nemesis, regular customer, etc.) which made up that person's social life. Let us say then that personal identity typically guarantees that one will have over the short term a unique *social continuer*. Whereas gross bodily continuity without pyschological continuity does not provide a social continuer, the massive psychological continuity described in the first presentation of Williams's case is plausibly taken to provide a social continuer, especially when the bodies involved are alike. So a tendency to trace *A* through a body swap rather than in accord with continuity of bodily life represents a fixation on the important *practical* concomitants of personal identity. Reinforcing the psychological-continuer effect we have the social-continuer effect, i.e., the tendency to trace a person in terms of his social continuer. This too will be a potentially distorting effect so long as we are not essentially occupants of complexes of social roles. The wide psychological view is thus parasitic upon the psychological- and social-continuer effects; it gives roughly correct conditions for the persistence of "personas"—psychological and social continuers of developed persons—but persons antedate, outlive, and may sometimes be outlived by their personas.

Secondly, if there is such a thing as the pure or merely determinable concept of personal identity, the concept of a persisting person as some or other unspecified kind of persisting locus of reflective mental life, and if the psychological-continuer effect leads us to trace such loci along lines of psychological continuity in cases in which bodily and psychological continuity come apart, then it is a short step to a more or less inchoate conception of ourselves as souls, i.e., primarily psychological and so not essentially physical loci of reflective mental life. That is, some of the primary intuitions supporting the wide psychological view may well depend upon a residual ten-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

dency to trace ourselves as nonphysical souls, a tendency which survives conversion to a secular world view, since it is at least partly born of the interaction between the pure concept of personal identity and the psychological-continuer effect, a tendency which should be at least mildly embarrassing to those physicalists who have advanced the wide psychological view.

None of this is to give a direct argument against the wide psychological view, but only to point out why the arguments for it should not carry conviction. There may be more direct arguments against the wide psychological view, but if they are needed they are best left for another time.[20] For to give them here would be to distract from the upshot of the present discussion, namely that the wide psychological view relies for its appeal on an analytical method which we have no reason to respect and indeed upon a method which if thoroughly pursued delivers the bare-locus view instead. If that is right, it is hard to see how the wide psychological view could now be rehabilitated. Indeed, it is hard to see why one would want to rehabilitate it, given the availability of the view that we are essentially human beings. For if we take ourselves to be essentially human beings we are able to locate ourselves in a broadly naturalistic conception of the world and find nothing problematic in our everyday practice of reidentifying ourselves and others on the strength of the continuous mental and physical functioning of the human organisms we encounter. Weighed against these advantages we have no real costs, but merely the deliverances of unconstrained imaginative conceits, deliverances which have misled the orthodox as a result of their uncritical overreliance on the method of cases, deliverances which actually show only that the concept of personal identity common to the adherents of particular and divergent conceptions of personal identity is just the unilluminating because merely determinable concept of some or other unspecified kind of locus of mental life.

MARK JOHNSTON

Princeton University

[20] See "Reasons and Reductionism," *op. cit.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE DEAD*

> His soul had approached that region where dwell the vast hosts of the dead. He was conscious of, but could not apprehend, their wayward and flickering existence.
>
> James Joyce, *Dubliners:* "The Dead"

DEATH opens the door to metaphysics. We must be bold and step through it. Plato has already traced for us, in the *Phaedo,* some of the intricate strands that tie together death and philosophy, the master insisting even that philosophy is itself a kind of cultivation of dying. There are, indeed, at least three topics that an analysis of death can be expected to shed light on: (1) through the meaning of death, the meaning of life, (2) the nature of nonexistence, and (3) the relation of living to temporality. By (1) I have in mind the Ecclesiastes problem: How can one's life have meaning if after all is said and done one is simply dead? "Seeing that which now is in the days to come shall all be forgotten, and how dieth the wise man? as the fool." The second problem concerns the fact that the dead no longer exist, and mere thought about the nonexistent (never mind what the appropriate attitude to adopt toward them may be) has been a thorn in the side of philosophers at least since the time of Parmenides, the Prince of Being. Finally, one can hardly expect to unravel the mystery of death without coming to grips with the even greater mystery of time—since it is, after all, the past that receives the dead and the future that holds the clue to our own mortality. More specifically, there is the problem, associated with Lucretius, of why we fear death but not the seemingly symmetric phenomenon of our prenatal nonexistence.

In what follows I will attempt to shed some light on the second and third problems. Thomas Nagel, in "The Absurd,"[1] has offered some interesting suggestions concerning the first, but this will not be our focus here. I will, however, be concerned to defend the following three theses: (1) Death is an evil, a misfortune, and one that befalls the nonexistent themselves. (2) The dead, appearances to the contrary, are not nothing. (3) Most people will never exist. The significance of these sobering facts is something that I hope will emerge as a consequence of the ensuing discussion.

* This paper was read at Philosophy Colloquia at Columbia University, Rice University, and Amherst College. I am grateful to these audiences for the healthy skepticism of their responses.

[1] In *Mortal Questions* (New York: Cambridge University Press, 1980).

0022-362X/87/8402/0084$01.80 © 1987 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

To begin, then, with the problem of nonexistence, note that we are naturally inclined to feel sorrow for the very person who has died, to continue to talk about him (or her), and to continue to adopt attitudes such as love and honor toward him. Our hearts here see more clearly than our heads; for these natural attitudes are disturbed by a beleaguering metaphysical conscience. How can one even assert, for example, that Socrates is dead? If this is a truth, whom is it supposed to be a truth about? Socrates? But there is, there exists, no such person. For, precisely because he is dead, he is no longer, and therefore there is no one of whom we may truly predicate death, or nonexistence. Further, how could death be an evil for Socrates? Nonexistent objects cannot feel anything; so how can they be expected to suffer from any evil, including the evil of death? Finally, how can one continue to love Socrates? There is no one there to love. There is, of course, in W. V. Quine's sense,[2] a "notional" kind of love, which we can feel even for Pegasus. But there is also, clearly, a "relational sense" of "love" according to which it is the object itself that we love; and how could one be expected to maintain this relationship when one of the terms is missing?

There are two prominent responses to these problems, both of which, I shall suggest, are misguided: (a) skepticism and paraphrase, and (b) what I will call *spatialization*. Before discussing these approaches, however, it is well to make certain that we have a clear concept of death itself. By 'death' I will mean precisely: "permanent postnatal nonexistence." By this definition, however, I do not mean to settle either the linguistic issue of what the English word "really means" or the factual question of whether, in this sense, anyone ever really dies. If Plato's suspicion (in the *Phaedo*) that the soul is im-

[2] "Quantifiers and Propositional Attitudes," in *The Ways of Paradox* (Cambridge, Mass.: Harvard, 1979). One should not underestimate the importance of such notional attitudes. As Sidney Morgenbesser and Mary Mothersill have pointed out, one feels a definite sorrow at the death of (the fictional) Anna Karenina. Clearly, one can appreciate in Anna Karenina traits that real people have, or can have, and we can, while reading the novel, momentarily suspend our realization that the characters are merely fictional. This accounts for the fact that, as Mothersill reminds us, one takes, as it were, a kind of pleasure in our sorrow for Anna, and that we recover from our grief with remarkable ease. Such attitudes should not tempt us, therefore, toward a theory that attributes to fictional characters the property of being real people, like us; we would, indeed, if this were done, lose the precise difference (as well as similarity) that exists between our irretrievable sorrow at the death of Jesus or Socrates, and our attitude to Anna Karenina. (Note, further, that, as Charles Parsons has remarked, the dead of the remote past play a role in our lives not unlike that of fictional characters. The mists of time lay an epistemological mantle over the distant dead, and we scarcely know whether we are grieving over fiction or reality.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

mortal is correct and if he is correct in suggesting that I am to be identified with my soul, then our discussion of death here will at least indicate exactly what misfortune it is that Plato has shown us we have been spared.

With these preliminaries out of the way then, what is the first approach to our problem? The paraphrastic refuses to take our attitudes to the dead at face value. To say that Socrates is dead is to indicate that the name 'Socrates' no longer has a bearer.[3] Those who pity Socrates are wasting their sorrow on one who suffers no evil; better to extend our pity to those who can suffer—the bereaved survivors. And to say that we still cling to Socrates himself, goes this way of thinking, is to speak loosely: we cling to all we have left, namely, our memories of Socrates, our impressions of him, and the impressions he has left on others.

I confess to finding this line of thinking shallow and unconvincing. To start with, the intuitions it challenges are clear, unequivocal, and quite literal: talk of the dead is of an entirely different order from, say, talk of the tooth fairy, which cries out for paraphrase. There is, moreover, a form of argument associated with Plato's *Euthyphro* which suggests the error in the paraphrastic approach. Surely Plato was saddened *because of* Socrates' misfortune, not: Socrates' death was a misfortune *because* of the ensuing sadness of Plato. The grievers grieve over the misfortune of death; death is not a misfortune because it gives rise to so many unhappy grievers.

Indeed, we find here a clue to one of our puzzles: death, as Thomas Nagel points out (in his seminal essay, "Death"; see footnote 1, above) like many other evils, is intrinsically relational. The lover betrayed by a lover has suffered a harm, even when ignorant of the harm done. To apply the *Euthyphro* again: the lover is grieved *because* he discovers the injury done him; not: the lover is injured *because* he becomes grieved over his discovery. The grief is due to the wrong, not the wrong to the grief. With purely relational evils, then, we are harmed because of what happens in relation to us, not because of a felt condition of our present state. Death, clearly, will be

[3] It was of course Bertrand Russell who announced, in "On Denoting," that his Theory of Descriptions enabled one to paraphrase away all apparent reference to the nonexistent. But his claim that descriptive phrases never refer, but at best "denote," does not survive scrutiny. [See my "Russell and Kaplan on Denoting," *Philosophy and Phenomenological Research*, XLVI, 2 (December 1985).] Moreover, it is surely too much to expect one's ontology to flow from one's semantics; the question of *what is* is not settled by the truth about 'the'. (Russell did have internal objections to the theory of nonexistents. But on these matters see below.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

an evil of this sort. But what is it the dead relate to, which explains their misfortune? Surely it is precisely the life they have been denied —the possibility of enjoying all that life is a precondition for. But surely this is too much! Is one really expected to swallow an account of death as a purely relational phenomenon between an object that *no longer* exists and a life that *never will* exist? In fact, yes; but the elucidation of these metaphysical mysteries will have to wait for later.

Nagel, as we have seen, recognizes the relational character of the evil of death, but his insufficient attention to metaphysics weakens his final account. Who is it, after all, that suffers from this (relational) evil? "[I]f there is a loss, someone must suffer it, and he must have existence" (*op.cit.*, 7). But the dead, alas, do not have existence, and so this account is as yet incomplete. Further, Nagel is led by his theory to deny to the unborn the right to suffer the evil of nonexistence. Whether or not he is right about this point is something we will come to shortly, but to complete his intuitions about the misfortunes of the dead it may become overwhelmingly tempting to try to find a way out by which the dead really can be seen to exist. If we could do this, Nagel's approach could be saved, and without any excessive metaphysical cost. This is indeed precisely what Harry Silverstein tries to do in "The Evil of Death,"[4] where he succumbs to temptation and maintains that Socrates, though dead, exists.

Silverstein accomplishes this feat by what I call the "spatialization" of time: treating time as if it were a kind of mere fourth dimension of space, he takes location in time to be as ontologically neutral as location in space. In this scheme of things your existence in the past tells us *where* (in time) you exist, not *whether* you exist (now—i.e., at all), just as to exist in New Jersey tells us where you are (in space) and not whether you are.[5] To exist at all, then, is to exist somewhere and somewhen. Silverstein gathers this conception from the weighty authority of Quine (with the "tenseless" existence of the quantifier) and Einstein (with the static geometry of the 4-dimensional space-time continuum).

Now I must say that I find this approach to Socrates' death just as unpalatable as its predecessor. My objections to it are as follows: (1) I find it exceedingly difficult to give up my intuition that dead people

[4] This JOURNAL, LXXVII, 7 (July 1980): 401–423.

[5] Arthur Prior had no use for this conception of the past as a mere place where the dead enjoy full ontological privileges. See "The Notion of the Present," in J. T. Fraser *et al.*, eds., *The Study of Time* (first conference; New York: Springer, 1972). Prior extended his displeasure to Special Relativity, if it is taken to imply this repugnant doctrine.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

simply do not exist. ([F]or living things it is living that is existing." [Aristotle, *De Anima,* 415b12].) And I do not mean merely that the dead do not *now* exist; for objects in time, what does not exist now does not exist at all. Since I find that my intuition that Socrates, as dead, no longer exists, is unshakable, I am more inclined to reject any philosophy that rejects this intuition than to give up the intuition itself. (2) The dead are not the only nonexistent objects that we seem compelled to admit. There are also, I will urge, (a) the unborn, (b) the (merely) possible worlds, and (c) the past and future themselves. How should we analyze these apparent nonexistents? The temptation may again be to spatialize—and this is, in effect, exactly what David Lewis has achieved with his possible-world theory. However, although the reasons for spatialization here closely resemble those in regard to time, philosophers have been (justifiably) wary of following Lewis in his extremely unintuitive spatialization of modality, according to which to exist in a possible world is really to exist. (This is not to say that anyone has successfully refuted Lewis—a point I have argued elsewhere.[6]) But then we must find some ad hoc device to account for these apparent nonexistents. Better to find a theory that will maximize our analysis of the nonexistent. (3) Finally, we should recall that Special Relativity is a mere empirical theory. It might have been false. If the four-dimensional Einstein-Minkowski view had been false of the actual world, would not death still have posed its problems? Is death not a problem already for Newton? Further, is there not something exceedingly paradoxical in the claim that certain experimental results, dealing with principles of symmetry in physics and results about the speed of light, should bear on the question of whether Socrates, as dead, really exists? Moreover, there are alternative philosophical accounts of the interpretation of Special Relativity (see, for example, the work of Milič Čapek,[7] Lawrence Sklar,[8] and Karl Popper[9]), especially as regards the verificationism that seems to imbue standard reconstructions of the theory. Thus (for reasons 1–3) Silverstein's spatialization approach is

[6] "On Time and Actuality: The Dilemma of Privileged Position," forthcoming in *British Journal for the Philosophy of Science.* Lewis's views can be found in his *Counterfactuals* (Cambridge, Mass.: Harvard, 1973).

[7] "The Inclusion of Temporal Becoming in the Physical World," in Čapek, ed., *The Concepts of Space and Time* (Boston: D. Reidel, 1976).

[8] "Time, Reality, and Relativity," in R. Healey, ed., *Reduction, Time and Reality:* (New York: Cambridge, 1981). Čapek seems to accept SR, but questions the excessive Eleaticism of the standard interpretations. Sklar discerns, at the heart of SR itself, a verificationist slide on which there seems no natural stopping place, short of a kind of "point-solipsism."

[9] *The Open Universe* (Totowa, N.J.: Rowman & Littlefield, 1982).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

not satisfactory. The clue to explaining the correct subject of the evil of death lies elsewhere.

The approach I favor is what the facts themselves seem to be urging us to accept: that death concerns the nonexistent as well as the existent; that, therefore, some things don't exist—i.e., that there are nonexistent objects. We should distinguish, therefore, between being something, being an *object* (which we express by 'there is', or '$(\exists x)$'), and being an *existing* object. Existence is that property, delicate as an eyelid, which separates the living from the dead. I express it with the predicate 'E!'. When I say "Some objects do not exist," I write it as follows: $(\exists x)\ (\sim E!x)$. I thus preserve existential generalization: from 'Fa' I can always conclude '$(\exists x)\ (Fx)$', and thus from "Socrates is dead" I conclude: "There is an object that is dead and this object is Socrates." That one can adopt this practice has been shown recently by the neo-Meinongian logicians. Here I follow Terence Parsons' notation in *Nonexistent Objects.*[10] I can agree, therefore, with Wittgenstein that Socrates' death is not an event in his *life;* but it is, for all that, a genuine event, and it does befall *Socrates.*

I believe, indeed, that talk of the dead, not talk of fiction, is the proper introduction to nonexistents. In discussing Socrates, for example, we clearly have a particular object in mind to refer to. He is a perfectly specific object. Having admitted this, we remind ourselves that he is, after all, dead, and hence no longer exists. "Ah," comes the reply, "but he used to exist." Precisely, and, therefore, since he no longer exists, he does not exist at all! (How easily one is tempted here to deny where this argument leads us, and to suddenly "spatialize" time, pretending that "past existence" is a kind of existence, "located" in the past.) Thus we are led to the conclusion that we have been referring all along to a specific nonexistent object and that to be is to be an object, whereas to exist is to be an existent object. Don't say: "But if we're referring to Socrates he *must* exist," nor: "Since he doesn't exist, we *can't* be referring to him." The facts are in front of us: we *are* referring to him, and he really *doesn't* exist. There simply is such a person as Socrates, and nothing, not even his

[10] New Haven, Conn.: Yale, 1980. I have noted among some philosophers the feeling that the theory of nonexistents, even if logically impeccable, is not to be taken seriously from a philosophical point of view—that it is, as it were, a mere metaphysical daydream. This attitude compares with that of sober theorizers in the past toward the relation of the new non-Euclidean geometries to an account of physical reality. But, just as Einstein's physics has shown that these new geometries are more than mathematical toys, we will come to see, I believe, that the theory of nonexistents is a tool for serious metaphysics and not a mere plaything for bored logicians.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

death, can ever erase this fact. (If the bad news is that you're going to die, the good news is that you won't "disappear"—i.e., become nothing.) Of course, had he never come to exist perhaps we would never have known him, nor been able to single him out and refer to him (but see below). But these semantic/epistemic facts do not touch the question of ontology. For surely we don't want to say that if Socrates had never existed there would never have been such a person as Socrates to come to exist (which is not to be confused with: "If Socrates had never existed there would have *existed* no such person as Socrates").

Now the plausibility of this approach turns on our ability to make sense of the quantifier, '$(\exists x)$', which no longer expresses existence, but rather "being" (which is *not* to be considered a kind of *existence*). Parsons does little to explicate this notion, and he could hardly be expected to define it, any more than classical logicians could define existence. Clearly, we are dealing here with logically primitive notions. Nevertheless, one can go some way toward explicating this new quantifier of 'being'.[11] It can be read: 'there is', or

[11] Charles Parsons sheds some light on this distinction in his important study, "Objects and Logic," *The Monist,* LXV (1982): 491–516, especially sec. 6, Being and Existence. He, too, invites us to consider the quantifier as expressing not existence but rather "being," and relegates existence to a special predicate. I find useful his comparison of the being/existence distinction with the Kantian-Fregean contrast of existence with actuality [*Wirklichkeit*]. Moreover, I agree with him that we must distinguish the concept of objects-in-general from the concept of existent objects, and that "if we use the devices of singular terms, predication, identity and quantifiers to make serious statements, then we are speaking about genuine objects" (506). Further, just as we refuse naively to admit objecthood to the referent of just any singular term (see text below), Parsons notes that, if our pure existential quantifier of "being" is to be applicable to abstract mathematical objects, it must be more restricted than Meinong's (since, for example, '$\sqrt{4}$' denotes a mathematical object, but 'the largest integer' does not). But, where Parsons seems to express sympathy with what he sees as Kant's view, that mathematical objects are "beyond possibility and actuality," I find my instincts to be more Platonic. I am more inclined to believe that, for such objects (as, perhaps, for God and the soul), their possibility implies their actuality. (See *Phaedo,* 105A–107B, as well as our text, below, on the ontological argument.) It is an interesting question, indeed, whether anything that lacks "matter" can ever be merely potential. Aristotle, for example, in *Metaphysics* Λ, concludes that the "unmoved mover," lacking matter, must be pure actuality. But how can what lacks matter be actual, active, (causally) effective? For Aristotle, the answer is: the unmoved mover is a "final cause," inspiring lesser beings to imitate its divinity. Similarly, Frege explains the peculiar "actuality" of the Platonic realm of pure thought: "Thoughts are not wholly unactual . . . their action is brought about by a performance of the thinker [who 'grasps' them] . . . and yet the thinker does not create them but must take them as they are" ["The Thought," *Logical Investigations* (Oxford: Basil Blackwell, 1977), pp. 29/30, with modifications in translation by Parsons]. Indeed, 'unmoved mover' is an excellent way to describe a "Platonic" object—a genuine yet "nonactual" object capable of being

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sometimes, 'there is possible', or 'there was' or 'there will be'. One should not exaggerate, however, our unfamiliarity with nonexistent objects that "are," or have "being." Just as we are intimate with various existent objects, we, most of us, are familiar with specific past objects (those we know who have died), and we are all used to dealing with the future. If we but let our eyes be opened, we will see that we have all along been trafficking with the nonexistent, as with the existent. The sense of our new quantifier is no more (and no less) mysterious than that of 'is possible' or 'did exist' (or 'died'). And by extending our ontology to the nonexistent, by embracing all of being, we are able to purchase the realm of the dead, the unborn, the future, and the possible, without resorting to either paraphrase or spatialization. We can thus preserve our most primitive intuitions about these domains.

The dead, then, we can now see, are not nothing. But what are they being contrasted with? What is nothing? (If I may be allowed to put it thus.) A partial answer is: Pegasus is nothing—it is nothing at all, since there is no such thing as Pegasus. My ontology is thus more Parmenidean than Meinongian. Names in fiction (and myth), I believe, are not genuine names for specific nonexistents, but "mock-names" with which we pretend to refer to existent objects. (There is no easy road to metaphysics; certainly, semantic naïveté is no substitute for ontological insight.) I thus agree with Keith Donnellan[12] and Saul Kripke[13] that to speak of Pegasus is to speak of nothing; and that is why, as Kripke urged long ago (*op. cit.*) in a Parmenidean moment, Pegasus, being nothing, not only does not exist, but could not come to exist. (What was not there in the first place, cannot come to be there.) To say this, however, is not yet to assert that one could not coin some new names, and attempt to refer to the "incomplete" objects that Terence Parsons has theorized about, objects defined by arbitrary sets of "nuclear" (i.e., safe, non-paradox-generating) properties, not necessarily closed under implication. This is an open ques-

---

contemplated by us. [For a discussion of the role of "grasping" Platonic senses in Frege's philosophy, see my "Frege, Perry, and Demonstratives," *Canadian Journal of Philosophy,* XI, 4 (December 1982): 725–752, as well as "The Path Back to Frege," forthcoming in *Proceedings of the Aristotelian Society:* esp. fn 47.]

[12] "Speaking of Nothing," *Philosophical Review,* LXXXII, 1 (January 1974): 3–32.

[13] "Naming and Necessity," in Donald Davidson and Gilbert Harman, eds., *Semantics of Natural Language* (Boston: Reidel, 1972); republished as *Naming and Necessity* (Cambridge, Mass.: Harvard, 1980); page references will be to the 1972 edition.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tion of metaphysics worthy of investigation. Indeed, this is just part of a fundamental question in metaphysics, to wit: Which essences define possible objects?

Progress on this problem would help us gain some insight into an important special case, in which the essence in question is God's and the subject is the ontological argument. It is worth while to linger on this issue for a few moments. Assuming, as is reasonable, that God's essence, as the greatest possible being, includes necessary existence, the question arises whether such an essence defines a possible being. Instead of investigating this pressing concern, however, too many philosophers have focused their attention on addressing "fallacies" in the argument—such as that some of its proponents have assumed too quickly that one can know that God is possible simply because we know he has an essence. And Kant, and others, have denied that existence is a property, or predicate, that can be included in an essence. But we have already seen that the mere advent of the existential quantifier by no means established Kant's thesis, since on our account we preserve both the quantifier and a predicate for existence. It is not a fallacy then, but an unresolved question of pure metaphysics whether or not a necessary being (such as God) is possible. And that it follows from the possibility of such a being that it does exist necessarily and, therefore, actually is, remains I take it, a valid conclusion, given a plausible modal logic, like C. I. Lewis's S5 (with Kripke's semantics),[14] where we have: $\Diamond\Box P \supset \Box P$ and also $\Box P \supset P$ (i.e., if possibly necessarily P, then necessarily P; and if necessarily P, then P).[15] There are, then, some important open

[14] "Semantical Considerations on Modal Logic," *Acta Philosophica Fennica*, XVI (1963): 83–94.

[15] There are those who question the importance of the ontological argument even if it is sound. They argue that: (1) they cannot recognize their God in the God of the argument, and (2) in particular they do not believe (their) God *necessarily* exists. I find this attitude puzzling in the extreme. Call their God "Jehovah," and that of our argument, "God." Now, first, I find it incredible that the Prophets would have considered themselves lucky that they found themselves in a world where Jehovah just happens to exist. At any rate, given that God and Jehovah both exist and are not identical, how can the pious not bow to God? For God, by definition, is perfect, and so it follows that Jehovah is not. Thus whatever divine attributes adorn Jehovah that make him worthy of worship, they are surpassed by God. Piety, I would think, compels us to worship not Jehovah, but God.

Before quitting this subject we should note an important lesson to be gathered from the ontological argument (as from the study of death): that what is possible is a genuine feature of the actual world, a feature that contributes to what is of value in it. What could be (or could have been) is as much a feature of our lives, as what is. This constitutes, I believe, the "romantic" conception of life, and it leads, inevitably, to a certain melancholy. For in this sublunar world of ours, nothing is ever all it could be. (This, I take it, is at the heart of Plato's philosophy.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

problems about which objects there are (are possible). But, I have been suggesting, not all beings, even the nonexistent, are as mysterious as God. The dead, for example, are a set of nonexistents easier to grasp than the unborn. We can name specific dead people and we know many detailed facts about them, whereas it is difficult to find a single unborn whom we can isolate and refer to with a name. This is no doubt part of why Thomas Nagel (*op. cit.*) and others are skeptical of the assertion that there are specific unborn, and, hence, that there is anyone there to suffer the deprivation of nonexistence.[16] Indeed, some have been moved to assert that concerning the merely possible there are only purely general facts. For myself, however, I find that this attitude comes dangerously close to the sin of conflating ontology with epistemology. The most that the above considerations show is that we cannot know, or refer to, specific unborn. (Though there may be certain exceptions. Given a suitable theory of counterfac-

[16] Those readers who are skeptical of my Platonism (which is not to be confused with Lewis's spatialization) will find great comfort in Hidé Ishiguro's contrary approach, elegantly presented in "Possibility," *Proceedings of the Aristotelian Society,* suppl. vol. LIV (1980): 73–87. "I do not hold," Ishiguro says, "that the actual world is one among the possible worlds. Possible worlds are ideal entities." Again, "Possible worlds, then, . . . are, I suggest, abstract entities, partially described, whose existence is invoked in my modal theories as models of my imaginative thoughts." And, finally, "Since possible worlds and possible objects are constructs of thought, they are not the kind of things to which actual things and the actual world can stand in any relation." My reaction to this, in a nutshell, is that it conflates abstract existents with concrete nonexistents. Possible people, like the dead and the unborn, are not a peculiar kind of abstract existent, but rather a perfectly ordinary kind of concrete object like you and me, who merely have the bad luck not to enjoy existence. When I contemplate having a child I do not turn my thoughts away from real, concrete children to the realm of (existent) *abstracta;* the (possible) children I am thinking of are of the ordinary garden variety, and really don't exist. What separates actual, living people from merely possible people is precisely their existence. As Kant's famous remark has it, "the real contains no more than the merely possible. A hundred real dollars do not contain the least coin more than a hundred possible dollars" (*Critique of Pure Reason,* A599 B627). Curiously, Ishiguro cites this very dictum of Kant's, in disapproval: "As Brentano says, possible dollars are not sums of dollars at all. They are not a kind of dollar—as a 1979 Hong Kong dollar or an 1800 U.S. dollar would be. Similarly, a possible world is not a kind of world like a chaotic world or a far-away world." As I read him, Kant seems to be making just the opposite point here. It is not Kant but rather Ishiguro who maintains that possible people are a "special kind of object." For it is she, not Kant, who claims that possible persons are *abstract, ideal constructs of our thoughts.* Kant, contrariwise, says that merely possible people are *just like you and me*—except for their nonexistence; and he is at pains to insist that the existent and the nonexistent do not constitute "special kinds" of object. Ishiguro may, of course, be correct that we sometimes think about possible people *by* mentally manipulating more accessible ideal constructs that stand in their place. But this only serves to highlight the fact that our intentional objects, i.e., the targets of our thoughts, are not the symbolic middle-men, but rather the possible persons themselves.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tuals, and the appropriate view of determinism, perhaps after an abortion we could name the very would-be person, say, "Newman," and thus refer to him.) But why assume that there are any unborn in the first place?

The point is that, once we admit the category of the nonexistent, we see that, as Plato notes in the *Phaedo,* the existent "come from" the nonexistent. Now, I find it exceedingly implausible to deny either that: (a) more people always could come to be born than already exist, or that (b) we could never "run out" of persons to be born. The "supply" of souls is unchanging; clearly, it must also be inexhaustible [what would it be to "run out" of (possible) people to be born?]. In fact, their number is awesome, and that explains the assertion with which this paper began, that: most people will never exist.[17] Indeed, from the point of view of pure ontology, the realm of existence is the merest dot of an "i" in the vast sea of being. All the more reason not to chauvinistically narrow our metaphysical preoccupations to what exists.

We are now in a position to address the question that Nagel was forced to dismiss: whether the unborn suffer from the evil of nonexistence. Assuming that life, as such, is a good (which is not to say that more life is always better than less), my answer, then, is: yes. This leaves open the question of whether the deprivation of nonexistence endured by the unborn is as great as that suffered by the dead, and

[17] It may be thought that there are in fact too many nonexistents—that Terence Parsons' theory is beset with Cantorian cardinality problems. George Bealer's incisive review, *The Journal of Symbolic Logic,* LIX, 2 (June 1984): 652–655, takes advantage of Parsons' generous formalistic schema for objecthood: "For any set of [nuclear] properties there is an object having exactly those properties," to raise threats from diagonalization. I am inclined to believe, however, that these problems concern not the nonexistents themselves but rather the abstract schematizations of Parsons. When similar crises befell set theory Kurt Gödel urged that we not be hoodwinked by shadows into ignoring what remains in the light: "Despite their remoteness from sense experience, we do have something like a perception of the objects of set theory. . . . The set theoretical paradoxes are hardly any more troublesome for mathematics than deceptions of the senses are for physics." [" What Is Cantor's Continuum Problem?" in Paul Benacerraf and Hilary Putnam, eds., *Philosophy of Mathematics,* 2nd ed. (New York: Cambridge, 1983), pp. 483/4).] Moreover, my concern in this essay is not with recondite metaphysical curiosities like Pegasus and the man in the moon (which are highlighted in Parsons' theory). Though my intuition may indeed falter in regard to these wayward and flickering objects, this hardly inclines me to doubt the reality of my deceased parents. Our intuition of the dead arises not from reflection on language or logical form, but rather from more direct epistemic contact: we know them in the same way we know many of the living, having at some time made their acquaintance. For some sensible advice about not letting our formalism set artificial limits to our ontological horizons, see Benacerraf, "Skolem and the Skeptic," *Proceedings of the Aristotelian Society,* suppl. vol. LIX (1985): 85–116.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

here the issue becomes more complex. That nonexistence is a misfortune at all for the unborn should by now no longer seem paradoxical. If one objects that the unborn *are* nothing, we respond that this is to confuse that which lacks existence with what really is nothing—as egregious an error as to conflate Socrates with Pegasus. If the reply comes that, nevertheless, the unborn *feel* nothing, then we remind ourselves that nonexistence, like death and betrayal, is a purely relational evil that requires no inner sensation to be endured. This still leaves us, however, with the issue of the comparative weight of the evil of nonexistence. Here, we may develop a suggestive conception of Nagel's (*op. cit.*). Certain misfortunes are not merely relational, but require a certain pattern in one's life. Thus, someone who, as an adult, has a terrible accident and is reduced to a state of total dependency, has suffered a misfortune, even though there is nothing intrinsically unfortunate about his final state (babies, for example, suffer no misfortune in enduring this state).

As I would put it, some evils are a function of the "curve," or "acceleration," of one's world-line, not of its properties at an instant. Thus, to "picture" certain misfortunes, one must display the curvature of one's world-line:

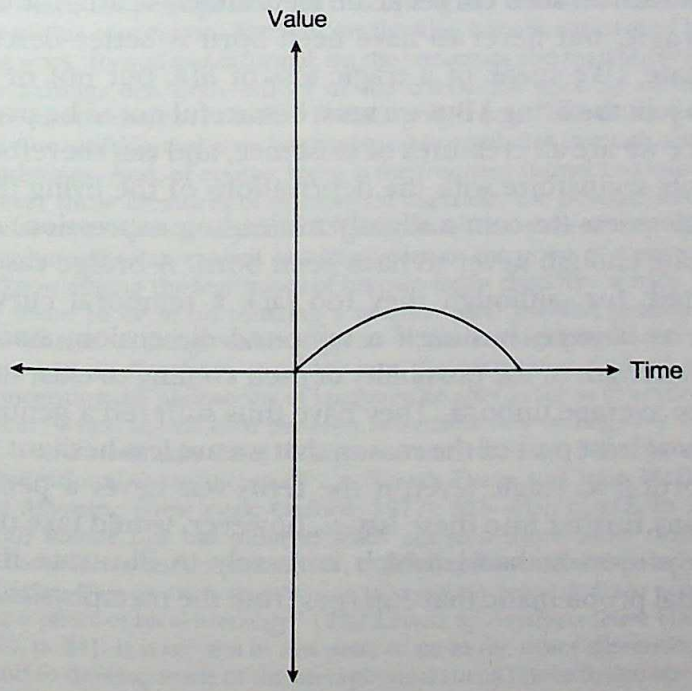


Figure (1)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Certain curves, it turns out (like the one described above), we take to be instances of misfortunes. Further, the curve will have to have a modal as well as a temporal dimension—possibilities that have been cut off from one's life. Thus, for example, if a woman with great potential as a violinist acquires a spouse who cuts off this career prematurely, she has (all else being equal) suffered a great harm, even though: (a) her "inner state," for all her married life, may never be unpleasant, and (b) the life denied her is a purely possible life that never came to exist.[18] Death itself, of course, would have to be analyzed by just such a "modal curve," except that in that case we would have a nonexistent person's relationship to a possible but nonactual continuation of his or her life. (For discussion of such 5-dimensional modal-Minkowski space-time diagrams, see my "On Time and Actuality," *op. cit.*) We can now approach our question. The obvious problem is how to compare a being (the unborn) that has a modal but no temporal curve with a being that has both (the dead): Is having no curve worse than having the worst curve? The poet seems to think so when he says in a similar context that " 'Tis better to have loved [lived] and lost than never to have loved [lived] at all." But, as the philosopher says, "It is proverbial that bards tell many a lie." The fact is that we want to use different categories of evaluation for those with (temporal) curves that indicate misfortune and those with no such curves at all: an untimely death, for example, can be *tragic,* but never to have been born is better described as unfortunate. (We speak of a tragic loss of life, but not of a tragic failure to join the living.) But we must be careful not to be prejudiced here, since we are all creatures of existence, and can therefore much more easily sympathize with the deprivations of the living than with the beinglessness (to coin a slightly misleading expression) of those unfortunate enough never to have been born. A bridge case here is the aborted; for, although they too lack a temporal curve, their modality, as it were, has itself a temporal dimension, since (for a time) the strength of the possibility of their coming to exist surpassed that of the average unborn. They have thus suffered a genuine *loss,* and that is at least part of the reason that we are less hesitant to speak of an abortion as tragic (even if the fetus was never a person). To proceed any further into these issues, however, would take us too far from the project at hand, which is merely to illustrate the fuller evaluational problematic that emerges from the metaphysical picture

[18] Thus Marlon Brando, in *On The Waterfront:* "I could have been a contender."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I have been suggesting. The silence of the dead is as nothing to the eternal quiet of the unborn, and I have thus tried to engage the reader on their behalf.[19]

The problem of the nonexistence of the dead thus beckons us along the path of metaphysics. But another mystery awaits us, since we have still to explore Lucretius' problem and the strange facts about time that it discloses. Why then do we fear our deaths, but not the seemingly indistinguishable period of nonexistence that preceded our births? We cannot, I suggest, grasp this problem if we insist on a spatialized conception of time, like that alluded to earlier, in connection with Silverstein. For if we treat time as a mere fourth spatial dimension, death and prenatal nonexistence are indeed mirror images of each other, and it would be irrational to fear the one and not the other. But if we don't adopt this conception of time, we open the door to the rationality of assuming an asymmetrical attitude here.

Let us follow John McTaggart and distinguish two fundamental features of time, the B-series and the A-series. The B-series characterizes times in terms of the (fixed and immutable) relation of "before and after," and is the subject of the mathematical physics of

[19] A few words are perhaps in order concerning the unfashionable metaphysical generosity of this conception. Much of my thinking here was provoked by Kripke's pioneering work, formal and informal, on the semantics and metaphysics of modality. In his writings one finds echoes of the traditional doctrine of essence and accident, of substances, not reducible to their properties, which retain their identity as they take on shifting and even unrecognizable attributes through temporal and modal wanderings. And, of course, there is the frequent formal and informal quantification over those creatures of ontological darkness, the possible worlds. But in spite of these suggestive conceptions Kripke himself has proved surprisingly shy of committing himself to an explicit ontology, even to the point of a suspect forbearance in acknowledging the legitimacy of his own brain child: the worlds. In point of fact there seems to be in his thinking a peculiar (and fruitful) tension between a formalistic scheme of things which suggests a classical metaphysic and a neo-Wittgensteinian urge to flee from ontology into the vernacular. Indeed in a related context concerning the philosophy of language he goes so far as to acknowledge this ambivalence: "I find myself torn between two conflicting feelings—a 'Chomskyan' feeling . . . and a contrary (late) 'Wittgensteinian' feeling" ["Is There a Problem about Substitutional Quantification?" in Gareth Evans and John McDowell, eds., *Truth and Meaning* (New York: Oxford, 1977): 325–419, p. 412, fn 56]. But the (ontological) silence this has induced leads not to a more sober metaphysic but rather to unexamined commitments; for, as Stanley Rosen observes in his commentary on Kripke, "we in turn must say that common sense is here a catch-phrase concealing a philosophical ideology" [*The Limits of Analysis* (New Haven, Conn.: Yale, 1980), p. 84]. It is my aim in this essay to go in the other direction, to bring to the light and so develop some of the metaphysical turns I have found unconcealed in the classical line in Kripke's philosophy.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

space-time (indeed Minkowski speaks of the *geometry* of the 4-dimensional space-time continuum). To recognize the B-series alone *is* to spatialize time. (Those who point out that the time variable in the definition of the interval in Special Relativity—viz. $\sqrt{x^2 - c^2t^2}$—has a different numerical sign than have the spatial dimensions, have hardly refuted my characterization. For this shows only that there are special *geometrical* constraints (with respect to time) on the space-time paths allowed by Special Relativity, not that we are not being offered a 4-dimensional geometrical space.) The A-series, on the other hand, characterizes times in terms of their shifting realizations as "now," or "past"/"present"/"future." The A-series gives us the time we experience and live in, and it carries with it some distinctive concomitants: (1) time passes, or "flows"—what was present is now past—and it flows toward the future; (2) the future, unlike the past and the present, is still open; and (3) location in the A-series is not ontologically neutral (to be in the past is not only not to exist now; for temporal objects it is not to exist at all).

Now, the Fundamental Problem of Time, I would suggest, is the relationship of the A-series to the B-series. Are both real, both objective, or is only one real, the other subjective? etc. What is disturbing is that the logic and metaphysics of the B-series have received increasing elucidation, especially since Einstein, but not only has the A-series not enjoyed such a renaissance, it has suffered from increasingly severe attacks from the friends of the B-series. These attacks come from logic, semantics, and physics. The attack from logic was launched by Aristotle. Intuitively, the past is now "necessary," in the sense of (now) unalterable (not because we cannot *get* there, but because it is already a fact, and you can't change the facts). Now for any seemingly future contingent proposition, P, by excluded middle: $P \vee \sim P$. But, however things turn out, either P or $\sim P$ will have been true in the past. Yet, since past truths are necessary, so is this past truth about the future. So the future, too, turns out to be necessary, contradicting condition (2) of the A-series.[20]

[20] If, with Aristotle, we reject such "logical determinism," the picture that results is one of a "moving," open future in which, so to speak, Reality itself "grows by accretion of facts" [Richard Jeffrey, "Coming True," in Cora Diamond and Jenny Teichman, eds., *Intention and Intentionality* (Ithaca, N.Y.: Cornell, 1979)]. Indeed, even Truth would have to expand, or Aristotle's ghost would return to haunt us. But how is this possible? Attempts to formalize this intuitive idea, even those as sophisticated as Richmond Thomason's model of "branching time" ["Indeterminist Time and Truth-value Gaps," *Theoria*, XXXVI, 3 (1970): 264–281], of which Jeffrey's seems to be a version, do not quite work [see my "On the Logic of Indeter-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The attack from semantics comes from the recent revolt against Frege, which led to the Kaplan/Perry theory of indexicals, like 'now' and 'present'. According to this theory, to call an event "now" is not to *say* anything about it—it is not to attribute any "property" or Fregean "sense" to it. Rather, the simple truth is given by the rule: at any time *t*, 'now' refers (simply, or "directly") to *t*. Since this rule obviously characterizes all times equally, no one time is ever "really" the one time that is now, or in the present: each time is now with respect to itself.[21] This leaves us with the argument from physics.

---

minist Time," this JOURNAL, LXXXII, 10 (October 1985): 548–559]. Worse, reflection suggests that no standard logicomathematical approach to the dynamic of time will succeed in capturing this feature: one cannot trap water with a net. Only a fool, it would seem, or a philosopher, would continue to pursue so elusive a quarry. But have I not overstated the recalcitrance of the problem? If the Big Bang Theory is correct, as well as General Relativity, then the Universe itself, i.e., space-time, has been "expanding" for eons. (Don't ask: where to? That would only betray mathematical innocence.) But an expanding universe is one thing, an "accumulation of Reality" (and thus of Truth) another. Mathematics can describe the former; the latter seems always to slip through its fingers.

[21] David Kaplan, "Demonstratives" (Ditto, Draft #2: UCLA, 1977); John Perry, "Frege on Demonstratives," *Philosophical Review*, LXXXVI, 4 (October 1977): 474–497. For a critique of this theory see my "Frege, Perry, and Demonstratives" (*op. cit.*) and "The Path Back to Frege" (*op. cit.*). In the first essay I drew attention to the fact that the ubiquitous Kaplan-Perry semantics for indexicals would lead inevitably to a spatialization of time (via 'now') and modality (via 'actual'—assuming the indexicality of actuality). Lewis was already peddling the latter, and in due course D. H. Mellor tried to sell us the former [see his *Real Time* (New York: Cambridge, 1981), where he wields Perry's semantics for 'now' as a club to beat down the A-series]. Now, although this is not the place to rehearse the detailed argument, we can illustrate at least briefly why Perry's theory does not get off the ground (and thus provide some semantic breathing room for the beleaguered A-series). The Kaplan-Perry rule for 'now', if put precisely, would be: "For all times *t* and speakers *s*, if *s* employs 'now' correctly at *t*, he refers to *t*." Now a rule is no good unless you can use it, but, if you try to employ this rule, it becomes obvious that, in grasping it, you get a handle not on any particular time, but only on a universal conditional on times (and speakers). The problem is that to use the rule to get to a time you must *instantiate* the universal quantifier, but, to accomplish this instantiation, you must already have a particular time *t* in mind. But how do you get to have it in mind? By describing it (e.g., as Saturday, 10:00 A.M.)? (This is vigorously denied by Kaplan and Perry.) By taking *t* to be the present moment—i.e., now? (This is circular; it is the rule itself that was supposed to show how we use 'now' to get to a particular time.) It seems, rather, that Kaplan and Perry have mistaken a necessary constraint on a mode of designation for a particular use of 'now' (that if 'now' is used at *t*, the mode of designation should determine *t*) for the mode of designation itself.

Curiously, Kaplan, in "Demonstratives," prides himself on his sharp separation of semantics from epistemology and insists (pp. 68/9) that his theory of "direct reference" needs no epistemological buttressing, like Russell's notion of acquaintance. Even Kripke, it will be remembered, had little use for the notion of a speaker's

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Given the relativity of simultaneity in Special Relativity, there is no unique set of events simultaneous with the present. Hence the now, the present, cannot but be a purely local phenomenon, not a global ontological one. These are serious challenges to the ontological credentials of the A-series. Can it survive them? This is not the place to settle the question. Given that the resolution of Lucretius' puzzle requires a solution to this dilemma, we can see why the problem of death is so profound. Without offering an analysis of this problem about time I will simply content myself with the extreme plausibility of the A-series, and for the sake of the present discussion assume its legitimacy.

It will be recalled, then, that an important aspect of the time of the A-series is that it is future-directed. We live in the "direction" of the future, and we attach different values to events, positive and negative, according to their position in our pasts or futures. Each of us knows, for example, the different attitude we would take if told we had had a terrifying toothache as a child, as opposed to being told that such a pain awaits us in ten minutes. In general, it is a fundamental feature of the human condition to focus our fear on future evils. If this is true for evils in general, how much more must it be so for the ultimate evil of death? It is thus precisely the futurity of death that renders it a more fearful evil than prenatal nonexistence. Moreover, we can see that, no matter how long we may extend our lives, we will not lessen the fear of death, since at the final moment one's death will still be "now." (Of course, unlike toothaches, we will never experience our own death—but by now this problem should not require any further attention.)

Note, further, that it is the mere futurity of death, coupled with the fact that we live in the direction of the future, which is relevant here, and not the openness of the future. For, to begin with, this is one aspect of our futures that is not open at all; as mortal creatures we are not free not to (eventually) die. Next, the fact that we may have some freedom in choosing when to die does not go to the heart

---

knowing who his referent is ("Naming and Necessity," *op. cit.*, p. 293). Thus it is that in post-Fregean semantics the competent speaker is relieved of the burden of knowing what he is talking about. Small wonder, then, that Hilary Putnam, always on the cutting edge, having contributed to the dissolution of the speaker-world relationship, has recently given up the world itself. For his version of Kantian (transcendental) idealism, see his *Reason, Truth, and History* (New York: Cambridge, 1981; see also "Reference and Truth," in *Realism and Reason* (New York: Cambridge, 1983), where he coins the phrase: 'existentially given' (p. 73), a marvelous neo-Kantianism in which ontology and epistemology are rolled into one.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of the matter, for the fear of death I am investigating here is not that of premature, or untimely, dying, but of the fact of (ultimate) death itself—i.e., of our mortality. One can cheat the angel of death with the small movement of a sharp blade, but one has not thereby tricked God out of our mortality. Nor will it avail to reverse this line of reasoning and argue from the very inevitability of death to the enormousness of its significance, since, of course, our prenatal nonexistence is every bit as inevitable.

The futurity of death, therefore, is its distinguishing feature, as opposed to prenatal nonexistence, and this is a feature invisible to the spatialized time of the B-series. Further, without the direction of time provided by the A-series, we will not be able to produce a full account of the "curves" of our lifelines discussed earlier. For it is essential to the evaluation of these "life-diagrams" that we note which direction is the future. A life, for example, that opens in misery and works its way to serenity is surely to be preferred to one that proceeds in the opposite direction. The evaluative study of such lifelines, then, cannot be reduced to a kind of pure geometry.

With these geometrical reflections, however, our study of the meaning of death draws to a close. Many questions, of course (especially of a valuational nature), remain open; my primary concern has been to provide a metaphysical framework with which to approach the dead. In attempting to unravel these mysteries we have pursued being into the nonexistent and come to grips with the most ephemeral and elusive of things: the flow of time. Being ourselves the future dead, it behooves us to engage (while we can) in serious philosophizing about death; and if we are thereby beckoned down the path of metaphysics we should not fear to tread it. It is indeed a narrow and treacherous way, but it is the only one there is.

PALLE YOURGRAU

Barnard College, Columbia University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

MORAL THEORY AND MORAL ALIENATION*

MOST moral theories share certain features in common with other theories. They consist of a set of propositions that are universal, general, and hence impartial. The propositions that constitute a typical moral theory are (1) *universal,* in that they apply to all subjects designated as within their scope. They are (2) *general,* in that they include no proper names or definite descriptions. They are therefore (3) *impartial,* in that they accord no special privilege to any particular agent's situation which cannot be justified under (2) and (3). These three features do not distinguish moral theories from other theories, nor indeed from most general categorical propositions we assert. Yet, in recent years, these features of moral theories have been the target of a certain concerted and sustained criticism, namely, that to be committed to such a moral theory, or to aspire to act in accordance with its requirements, results in what has come to be known as *moral alienation.*[1] Moral

* Work on this paper was supported by an Andrew Mellon Postdoctoral Fellowship at Stanford University, 1982–84. The paper is excerpted from Chapters VI and XII of a much longer discussion, *Rationality and the Structure of the Self* (in progress). Earlier versions have benefited greatly from criticisms by Akeel Bilgrami, Jeffrey Evans, and members of audiences at Wayne State University, Penn State, Georgetown, UC/San Diego, North Carolina State, Wesleyan, Memphis State, and the University of Minnesota, where this paper was presented in the winter of 1986.

[1] This term was coined by Bernard Williams. He elaborates the criticism in a series of papers: "A Critique of Utilitarianism," in J. J. C. Smart and Williams, *Utilitarianism: For and Against* (New York: Cambridge, 1973); "Morality and the Emotions," in his *Problems of the Self* (New York: Cambridge, 1973); "Persons, Character, and Morality," in *Moral Luck* (New York: Cambridge, 1981); and "Utilitarianism and Moral Self-indulgence," also in *Moral Luck*); and in his recent book, *Ethics and the Limits of Philosophy* (Cambridge, Mass.: Harvard, 1985). More recent proponents of Williams's criticism include Lawrence Blum, *Friendship, Altruism and Morality* (London: Routledge & Kegan Paul, 1980); Michael Stocker, "The Schizophrenia of Modern Ethical Theories," this JOURNAL, LXXIII, 14 (Aug. 12, 1976): 453–466; Michael Slote, *Goods and Virtues* (New York: Oxford, 1983); and Susan Wolf, "Moral Saints," this JOURNAL, LXXIX, 8 (August 1982): 419–438.

Williams's criticism is anticipated—and a rebuttal to it leveled—by John Rawls's observation that "in times of social doubt and loss of faith in long established values, there is a tendency to fall back on the virtues of integrity: truthfulness and sincerity, lucidity and commitment, or, as some say, authenticity. If no one knows what is true, at least we can make our beliefs our own in our own way and not adopt them as handed to us by others. If the traditional moral rules are no longer relevant and we cannot agree which ones should take their place, we can in any event decide with a clear head how we mean to act and stop pretending that somehow or other it is already decided for us and we must accept this or that authority. Now of course the virtues of integrity are virtues, and among the excellences of free persons. Yet while necessary, they are not sufficient; for their definition allows for most any content: a tyrant might display these attributes to a high degree, and by doing so exhibit a

0022-362X/87/8402/0102$01.70 © 1987 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

alienation, according to this criticism, consists in (i) viewing one's ground projects from an impersonal, "moral point of view" engendered by one's acceptance of the theory; (ii) being prepared to sacrifice these projects to the requirements of moral principle; and (iii) making such a sacrifice specifically and self-consciously in order to conform to these requirements.

Moral alienation is said to manifest itself in one (or both) of two ways, depending on the nature of the project thus susceptible to sacrifice. One may be alienated *from oneself,* if the project consists of tastes, convictions, or aspirations that are centrally definitive of one's self. In this case one's commitment to the project can be at best conditional on its congruence with one's moral theory. It is claimed that this must make for a rather tepid and unenthusiastic commitment indeed.[2] Alternatively, one may be alienated *from others,* if the project is an interpersonal relationship such as a friendship, marriage, or collegial relationship. In this case one's responses to the other are motivated by one's awareness of what one's moral theory requires. It is claimed that this obstructs a genuine and unmediated emotional response to the other as such.[3]

My aim here will be to argue that this very compelling criticism—call it the *moral-alienation criticism*—is nevertheless misdirected. The real culprit is not any particular moral theory, but rather a certain familiar personality type that may or may not adopt it.

I. THE MORAL POINT OF VIEW

Implicit in the above criticism of moral theory is a distaste for the point of view that our adoption of a moral theory purportedly forces on us. It is claimed that to take the moral point of view is to view ourselves and others *sub specie aeternitatis,* in abstraction from our personal circumstances and relations to others, and hence with a "detachment . . . from the level of all motivations and perceptions other than those of an impartial character."[4] The idea seems to be that, quite regardless of the particular *content* of a particular moral

certain charm, not deceiving himself by political pretenses and excuses of fortune. It is impossible to construct a moral view from these virtues alone; being virtues of form they are in a sense secondary" [*A Theory of Justice* (Cambridge, Mass.: Harvard, 1971), ch. IX, sec. 78, p. 519].

[2] This thesis, and, more generally, moral alienation from oneself, is a focus of Wolf's discussion (*op. cit.*, pp. 427–435).

[3] This claim, and, more generally, moral alienation from others, provides the target of Stocker's discussion (*op. cit.*), as well as of Blum's discussion (*op. cit.*, ch. III).

[4] This is Williams's characterization in "Persons, Character, and Morality," *op. cit.*, pp. 2–5; also see *Ethics and the Limits of Philosophy, op. cit.*, pp. 19/20, 51, 65–67, 103/4, 110, 118.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

theory, we are forced to view the world without reference to our place in it, simply in virtue of the theory's impartiality. Thus it is not a necessary condition of such a perspective that the moral theory in question identify impartiality as itself a value: it need not direct us, say, to treat all human agents as equals or to assign the happiness of each a numerical value of 1 and no more than 1. Instead, the prescription may be, for example, to cultivate one's talents and moral dispositions or to render aid to the needy when it does not involve undue personal sacrifice. In these cases, too, the impartiality of the principle requires the so-called "moral point of view": My reason for cultivating my talents and moral dispositions is not that I, Piper, have a special attachment to *my* own potential for creativity and moral virtue, but rather that it happens to be, in this particular case, my potential that stands in the relation "one's own" to me; hence it is my potential that the prescription enjoins me to cultivate. Similarly, my reason for helping the needy is not that I am moved by compassion for the needy whom I see before me, but rather that my moral theory prescribes that one render aid to the needy, and I identify these individuals before me as "needy." In all such cases, the impartiality of the moral principles I accept is supposed to entail that I view myself and others without any special reference or relation to my personal circumstances and relations.

This version of the moral point of view fails to distinguish between *impartiality* and *impersonality*. According to the *Oxford English Dictionary, impartial* means "unbiased, unprejudiced, just, fair, or equitable." This corresponds roughly to the characterization of impartiality given in the introductory remarks of this discussion. By contrast, *impersonal* means "having no personal reference or connection." This is the view of oneself and others *sub specie aeternitatis* on which the moral-alienation criticism seems to rely. Proponents of this criticism assume that there is an intimate connection between the impartiality of the principles of moral theory and the impersonality of the point of view one must take as a consequence of adopting it. But this assumption is false. One may adhere to impartial moral principles without adopting an impersonal point of view, and one may adopt an impersonal point of view without adhering to impartial moral principles.

Consider the first possibility: that an agent adheres to impartial moral principles without adopting an impersonal point of view. Suppose you are personally invested in a moral theory that contains various impartial prescriptions of fairness, sympathy, honesty, compassion, and so forth. Also suppose that there are clear psychological reasons for your personal investment, deeply rooted in your per-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sonal history. Perhaps, in addition to having had a sound moral upbringing, you discovered early on that these deeply instilled principles were your only resource for coming to terms psychologically with repeated personal injustice or confusing or threatening personal encounters. Suppose further, then, that your investment in these principles informs your entire social and personal life. You try to do what is right, to be fair and honest in your dealings, to understand and sympathize with others, and to respond to them compassionately and without prejudice, as your moral convictions prescribe. These principles also inform your private life: you attempt to secure and maintain good physical health, to live modestly but tastefully, and not to deceive yourself about who you are or to what you aspire. The moral prescriptions that guide this conduct are universal, in that they apply not only to you, but to all rational human agents. They are also general, in that they include no proper names or definite descriptions. They are therefore impartial, for they do not permit you to accord any special privilege to your personal requirements, merely in virtue of the fact that you are the agent whose behavior you are evaluating. Of course your moral theory includes provisions for different circumstances and social relations, for example, that the elderly deserve special respect for their wisdom and experience, that one has special obligations to family and loved ones, and so forth. Nevertheless, it applies universally, generally, and impartially, for there is nothing in it tailored to fit your particular situation.

Now if your adherence to this moral theory implied an impersonal point of view, then since this point of view is the symptom of moral alienation, it would follow from your overriding investment in this moral theory that you were alienated from those of your central desires and ground projects thus overridden. But surely you are so alienated only if your personal investment in your ground projects outweighs your personal investment in the moral theory with which they may conflict, and this is an open question. We cannot simply *assume* that you are morally alienated *merely* by virtue of your overriding personal investment in your moral theory. For that would make the moral-alienation criticism a definitional rather than an empirical thesis.

Let us say that an agent A is *personally invested* in some state of affairs x if (1) x's existence is a source of personal pleasure, satisfaction, or security to A; (2) x's nonexistence elicits feelings of dejection, deprivation, or anxiety from A; and (3) these feelings are to be explained by A's identification with x. A *identifies with* x if A is disposed to identify x *as* personally meaningful or valuable. According to the criterion of personal investment, you are alienated from

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

your ground projects if (because you identify with them) sacrificing the possibility of their realization when your moral view so prescribes elicits the feelings of profound loss described by (2). You are alienated from your moral theory, on the other hand, if, because you identify more completely with this theory, these feelings of loss result from pursuing your central projects at the expense of your moral theory. For in this case, it is your moral theory that is the focus of your personal investment, not your desires and ground projects.

Proponents of the moral-alienation criticism do not speak directly to the importance of something like the criterion of personal investment in determining when, whether, or from what a person is morally alienated. They seem, on the one hand, to incorporate it, by referring to that from which they presume agents to be alienated as "commitments" and as that with which they "identify." On the other hand, an important supposition of the moral-alienation criticism is that a morally alienated agent regards her ground projects with detachment—i.e., *without* these feelings of loss. But it does not follow from the fact that the abdication or sacrifice of certain things with which we identify fails to elicit profound feelings of loss from us that we are alienated from those things. For they may have been only peripherally important sources of self-identification to begin with.

Consider another, analogous conflict between impartial principles and psychologically central desires. Suppose you have a settled, long-standing, and recurring desire to smoke. You also hold wholeheartedly the universal, general, and impartial conviction that it is both unhealthy and inconsiderate to others to smoke. Although you are frequently tempted to smoke, the force of your conviction usually enables you to resist this temptation. Your unqualified personal investment in the view that smoking is a moral evil leads you to fear your desire to smoke and to anticipate its onset with anxiety at the possibility that you may give in to it. When it occurs, you are simultaneously torn by craving for a cigarette and by self-disgust at your inability to rid yourself of this craving once and for all. When the desire passes, you rejoice, relax, and hope that you have seen the last of it. Of course you experience temporary feelings of deprivation, dejection, and anxiety at not smoking when the desire to smoke overtakes you. But these feelings are to be explained by your physical addiction to smoking, not your identification with it. Though you often desire to smoke, you have no personal investment in smoking. If you have no personal investment in smoking, it is inappropriate to describe you as alienated from your desire to smoke, even though that desire may be long-standing, central, and powerful.

If one is not commonly assumed to be alienated from the desire to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

smoke when one withstands it in favor of the impartial prescription not to smoke, it is unclear why one should be assumed to be alienated from any other desires when one sacrifices them in favor of an impartial prescription to, say, render aid to the needy. This supposition of an agent's personal investment in her projects at the expense of her moral theory needs to be demonstrated, not taken for granted.

II. THE IMPERSONAL POINT OF VIEW

Next consider the possibility that an agent adheres to an impersonal point of view without adopting a set of impartial moral principles. Take the question of whether or not I should save my husband Jeff first from some natural disaster, and suppose my moral theory sufficiently fine-tuned to yield the answer that I should—say, by the inclusion of a special-obligations-to-loved-ones clause. The objection then would be that I am morally alienated from Jeff nevertheless, if I am motivated to save him first because my moral theory prescribes it, rather than out of love for him. The moral-alienation critic would say that my investment in this theory detaches me from my love for Jeff, since it is only in virtue of my theory that I am overridingly and unambivalently motivated to save him first. My impersonality is evinced by my primary attachment to my moral theory.

But the same objection could be made even if no such theory intervened between me and Jeff. For my desire to save him first may also intervene between Jeff and me. In this case, the complaint would be that my investment in the satisfaction of my desires—especially, let us suppose, the altruistic and other-directed ones—takes precedence over my love for Jeff, since it is only in virtue of my unsatisfied desire to save him first that I am overridingly and unambivalently motivated to do so. Here my impersonality, my lack of personal connectedness to Jeff, is evidenced not by my attachment to my moral theory, but rather by my attachment to my own desires.

In both cases, there *may* be some validity to these complaints. I may be, indeed, so enamored of my moral theory and the fact that I subscribe to it that I really do regard other moral agents as nothing more than occasions for instantiating its precepts. But alternatively, I may be so committed to the satisfaction of my other-directed desires that I regard other moral agents in a similarly superficial way, as mere occasions for exercising my beneficence. This is the stereotype of the "do-gooder," whose moral behavior somehow seems entirely self-aggrandizing and neither elicits nor presupposes any attachment to the individuals her actions serve. The do-gooder, too, foregoes the depth and insight that accompanies attention to the specifics of who they are, for the sake of an essentially egocentric conception of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

reality. In the first case, the self is personally invested in a theoretical moral stance that degenerates into impersonality because it is used to deflect and disguise unmediated interpersonal contact. In the second case, however, the self is invested in a *non*theoretical moral stance that also generates into impersonality because it deploys others as a means to the achievement of its personal moral ideal.

Thus I may subvert my personal connection with myself and others, not only by interposing an impartial moral theory between us. I may accomplish this by interposing self-aggrandizing emotions and desires as well. For in both of these cases, the very real problem to which the moral-alienation criticism calls attention is not just personal detachment, but a deeper, more generalized pathological narcissism, of which the condition identified as "moral alienation" is merely a contingent and localized symptom.

By *narcissism,* I shall mean that persisting state of the self characterized by an excessive preoccupation with one's self-image and image in the eyes of others, with one's personal flaws and assets and an unrealistic ideal of perfection against which they are measured, and with a self-oriented conception of one's relationship to others. Moreover, a self is narcissistic if it cannot tolerate others' independence of its expectations and requirements, nor appreciate their independence as intrinsically valuable. Finally, a self is narcissistic if these preoccupations effectively shield it against unmediated interpersonal vulnerability or contact, and against the trauma of personal growth that frequently results.[5] Among the manifestations of narcissism some psychologists have observed are envy and an inflated sense of one's own importance, the consequent devaluation of others and of their attentions to oneself, and the inability to empathize with others, to experience a sense of connectedness with them, and to form deep attachments to them. At the same time, narcissists are often highly dependent on others to buttress an extremely fragile and volatile self-esteem. They are beset by occasional eruptions of self-righteously sadistic anger and nightmares of self-contempt, both of which are recycled to fuel the grandiose belief that their ideals and aspirations are higher and purer than anyone else's. Consequently, narcissists are frequently smug, condescending, and seemingly remote as well.[6]

[5] This characterization is based on the criteria described in *DSM III: Diagnostic and Statistical Manual of Mental Disorders,* 3rd ed. (Washington, D.C.: The American Psychiatric Association, 1980), pp. 315–317.

[6] Dr. Otto Kernberg discussed these symptoms in a lecture at the University of California at San Francisco on May 2, 1984. Also see Kernberg's books. *Borderline Conditions and Pathological Narcissism* (New York: J. Aronson, 1975), part II;

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

It is not difficult to understand how a narcissistic self might have a greater allegiance to its favored moral theory than to other people, or, alternatively, a greater allegiance to its altruistic self-image than to the individuals thereby served. Nor is it difficult to understand in what sense a narcissistic self might view others from a detached or impersonal perspective, and how its concern with the opinions of others might be accompanied by an inability to establish genuine and unmediated contact with them. That is, it is not difficult to see how moral alienation might be a significant problem for a narcissistic self.

Let us speculate on the rationalized form such an attitude might take when the narcissistic self is confronted with evidence of its own narcissism, say, in the form of a complaint that one's aspiration to sainthood seems largely unaccompanied by any personal warmth. If narcissism functions as a defense against the intrusion of an undisguised other into the domain of the self, a natural response to such a complaint would be to denigrate personal warmth as an invasion of privacy. Thus, for example, it might be argued that attempting to be familiar or cozy with everyone one meets is the worst form of moral inauthenticity of all, that it is irrational and self-indulgent to allow others to make importunate emotional demands on one, or self-defeating to allow them to take advantage of one's generosity and moral concern, that it is delusory or even self-destructive to "spread oneself too thin," "try to save the world," or be a martyr, and that the preservation of a private realm within which the universalistic demands of morality cease to apply is a necessary condition for having any sustained moral impact whatsoever.

These claims derive their persuasiveness from the recognition that few of us indeed can be morally effective as martyrs. But, deployed as narcissistic defenses, they rationalize the detachment of the self from the moral requirements of others, and the withdrawal of the self into a private domain in which those demands can be safely disregarded. By contrast with views that justify the need for privacy and personal fulfillment as a condition of greater moral compassion and commit-

---

and *Severe Personality Disorders* (New Haven, Conn.: Yale, 1984), part III. For a related and insightful discussion that does not, however, use the concept of narcissism explicitly, see David Shapiro, *Autonomy and Rigid Character* (New York: Basic Books, 1979).

It is important to emphasize that the following discussion is intended to chart some of the connections between narcissism as I have defined it and the so-called "impersonal" point of view. It should *not* be taken to imply that actual proponents of the moral alienation criticism are narcissists, since of course it is the *philosophical* force of the thesis that has garnered it so many adherents. I am grateful to Jeffrey Evans for alerting me to this possible misreading of the argument.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ment, the narcissist would claim that there is a certain realm in which the requirements of moral compassion and commitment simply fail to apply. That is, a sanctuary for the individual self is not justified as a necessary condition of sustaining and strengthening its moral ties to others, but instead as a sufficient condition of sustaining and strengthening the self to withstand them. Others, in this view, are regarded as intrusive or disruptive of the equilibrium of the self or as exacting too great a demand on its resources, rather than as enrichening them. The narcissistic self is distinguished, then, by the subordination of its moral commitments to its need for eminent domain.

A narcissistic self also might be expected to regard any impartial moral theory of which it is a potential *object* as unsympathetic or distasteful. For, by definition, an impartial theory refuses to accord privileged or exceptional status to the requirements of any self, including narcissistic selves. But we have already seen that one of the defining features of the narcissistic self is the arrogation to itself of value and importance that others are perceived to lack. From the viewpoint of the narcissistic self, the privileged status of its particular requirements and its right to special treatment are both justified by their special and superior value in its own eyes.

Moreover, from the perspective of a narcissistic self committed to the preservation of its internal boundaries against unmediated contact with others that threaten or disrupt it, the obligations, for example, to treat all human beings fairly or not to be biased by one's personal preferences would naturally present themselves as particularly odious competitors to that commitment. For recall that another one of the defining features of such a self is that its central desires are specifically self-oriented, regardless of the content of those desires. The actual requirements of impartial morality disturb the integrity of the narcissistic self by threatening this orientation. For they demand, not merely the personal desire to conform to them, but rather an unmediated comprehension of and sympathy for the needs and requirements of others which, for a narcissistic self, are in direct competition with its own. We might expect, then, that the integration of personal needs and desires with the requirements of moral principle would be regarded by the narcissistic self as an abdication or sacrifice of selfhood, and rejected accordingly.

Not just moral alienation, but a more generalized social alienation is a predictable outcome for a narcissistic self. For if one's primary concern with others is the nature of their relation to oneself and if their behavior is invariably interpreted as evidence of this relation, then, obviously, one's view of others will be mediated by this inter-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pretation and, correspondingly, detached from their independent reality. In this case, whether the terms of this interpretation are theoretical or affective is largely irrelevant. An agent who saves her spouse first in order to fulfill her moral obligations or satisfy her benevolent desires is psychologically and morally crippled, but not because of her moral theory. She is crippled because her preoccupation with her own rectitude overrides the dispositions and behavior that her moral theory prescribes.

Thus the real problem to which the moral-alienation criticism importantly draws our attention does not lie with moral theory. For we have seen that one may adopt such a theory without assuming an impersonal point of view and that one may assume this point of view without adopting a moral theory. An impersonal point of view is instead symptomatic of a more generally corruptive and debilitating pathology, namely narcissism, which bears no necessary relation to moral theory at all.

III. MORAL INTEGRITY RECONSIDERED

My attempt to preserve the unsullied character of moral theory against the moral-alienation criticism may remain suspect because of one feature of that thesis to which I have so far paid little attention. This is the background Humean view, according to which all action must be motivated by desire, however weakly that notion is understood. From this vantage point, it may seem that any account of moral motivation to act on impartial principles of moral theory must either presuppose a desire to act on those principles—in which case my desire to save Jeff first, for example, is not unmediated after all—or else it can issue only from the impersonal point of view *of those principles themselves*. Either possibility would belie my contention that psychological narcissism has nothing to do with moral theory. In what follows, then, I want to limn a third possibility. This is that moral conduct is motivated directly and unmediatedly by those impartial moral principles that are partly and necessarily constitutive of the *personal* point of view.

Begin by reconsidering my decision to save Jeff first because my moral view condones it. Let us bite the bullet and describe this as a case of my being motivated to save Jeff first by respect for an impartial moral imperative, derived from that part of my moral theory which assigns me special obligations to loved ones, to aid loved ones first when rendering aid to the imperiled.[7] The complaint would be

[7] Clearly this formulation of the case is Kantian in spirit. However, I do not mean to ascribe to Kant the solution that follows, since I think it is only part of the story

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that I am then motivated by a desire to adhere to my moral theory, and not by my concern for Jeff.

But this does not follow. Recall first the distinction between a purpose and a motive for acting.[8] A *purpose* for acting is the goal, end, or intentional object to the achievement of which my behavior is directed. A *motive* for acting is the psychological cause of action, i.e., that which moves me to behave intentionally. Under the influence of the Humean conception of the self, many philosophers assume that the purpose of my action is necessarily its psychological cause as well. They assume this because they suppose that the purpose of my action must be the object of a desire or, minimally, of a "pro-attitude" toward it, which suffuses it with a weak but rosy glow and inspires me to pursue it.[9] According to one well-known account,[10] I have such a desire or pro-attitude toward this object if, when I fail to achieve it, I experience disappointment, frustration, or regret. But that my action is *directed toward* the achievement of this object does not imply, even minimally, any such *pro-attitude* toward it in any nontautological sense.[11] For example, I may be caused to purposefully peel the label off the ale bottle, not by any pro-attitude toward peeling the label off the ale bottle, but rather by anxiety, or habit, or the perception of the dampness of the bottle. If I am prevented from doing so, I may experience neither disappointment, nor frustration, nor regret. Hence that my action is *directed toward* the achievement of this object does not imply that it is this object that *causes* me to

---

for Kant himself. That story is developed in my "Kant's First- and Third-person Criteria of Humanity," and "Kant's Idea of Reason," unpublished papers, 1985.

[8] This distinction is first made explicitly by Kant, in *The Critique of Practical Reason,* Lewis White Beck, trans. (New York: Bobbs-Merrill, 1956); see Remark I to Theorem IV. H. A. Pritchard relies on this distinction, although he uses it to different ends, in "Does Moral Philosophy Rest on a Mistake?" *Mind,* XXI, 81 (January 1912), 21–37.

[9] See, for example, Donald Davidson, "Actions, Reasons, and Causes," this JOURNAL, LX, 23 (Nov. 7, 1963): 685–700; and Alvin Goldman, *A Theory of Human Action* (Englewood Cliffs, N.J.: Prentice-Hall, 1970), chs. III and IV. This assumption is central to Alan Gewirth's argument for the Principle of Generic Consistency in his vigorous and challenging *Reason and Morality* (Chicago: University Press, 1978), chs. 1.12–1.14, 2.1–2.6.

[10] Richard Brandt and Jaegwon Kim, "Wants as Explanations of Actions," this JOURNAL, LX, 15 (July 18, 1963): 425–435.

[11] And it is, of course, only the nontautological sense we are interested in, when addressing the question of what discrete, causally efficacious internal events move the agent to action. The tautological sense of 'desire' that makes it a conceptual truth that actions are motivated by desire gives at best a prematurely conversation-stopping answer to this question. This problem is discussed in greater detail in chs. IV and V of my *Rationality and the Structure of the Self, op. cit.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pursue it. Moreover, even if I did have a pro-attitude toward this object, even this fact would not imply that this pro-attitude is what causes me to pursue it: It is an open question whether it is my pro-attitude toward peeling the label off the ale bottle or my anxiety which causes me to do so. The purpose of an action need not supply its motive.

Of course some purposes of action *do* supply its motives, as when the intentional object at which my action is directed is one I desire, or aspire, or resolve to achieve. Desires, aspirations, and resolutions are psychological causes of action that take the agent's purposes as intentional objects and would not be motivationally effective without them. These are the cases in which it makes sense to describe the agent as having a motivationally effective "pro-attitude" toward the purpose of the action. I shall describe these causes of action as *forward-looking motives*. My claim is that not all action is caused by forward-looking motives.

For there are other psychological causes of action that are unrelated to the purpose of my action, and instead presuppose perceived intentional objects as causes. For example, perceived traffic jams cause frustration, which in turn motivates honking the horn. Honking the horn is a fully intentional action. I may have a pro-attitude toward it, but then again I may not. In either case, honking the horn need not be motivated by its purpose. Instead, it may be motivated by an emotion that is caused, in turn, by the perception of an intentional object. Call such psychological causes *backward-looking motives*. My claim is that much action is motivated solely by backward-looking motives.[12]

Backward-looking motives, in turn, may be of two kinds. In the example just described, the immediate psychological cause of action is an emotional reaction to a perceived intentional object. Describe such motivationally effective emotional reactions as *affectively motivating states*. Affectively motivating states constitute one kind of backward-looking motive. But, sometimes, perceived intentional objects can elicit a goal-directed behavioral response almost automatically, without the intervention of an affectively motivating state, if the disposition to respond to that perceived intentional object in that way is deeply instilled, as when I respond to the perceived ringing of

[12] My distinction between forward- and backward-looking motives parallels Michael Stocker's distinction between the "in order to"/"for the sake of" and "out of/from" locutions. See his "Values and Purposes: The Limits of Teleology and the Ends of Friendship," this JOURNAL, LXXVIII, 12 (December 1981): 747–765.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the telephone by picking it up and saying, "Hello?" Call these *perceptually motivating states*. In these cases, the mere perception of an object is motivationally effective in causing an overt behavioral response directed toward a different object. Backward-looking motives may be either affective or perceptual.

Now, according to the prevailing Humean model of motivation, any such backward-looking motive must be followed by a forward-looking motive, namely, a desire, if it is to cause action. Thus, for example, the Humean picture implies that my feeling of expansiveness, caused by my having just got a raise, can only indirectly cause me to scatter dollar bills in the street, by first engendering in me a *desire* to scatter dollar bills in the street. But no such desire (non-tautologically construed) is necessary to explain action. It is often sufficient that deeply inculcated norms of social behavior simply dispose me to react or behave in certain ways in response to my perception of a situation as being of a certain kind.[13] In the present example, my emotional reaction to getting a raise, i.e., my feeling of expansiveness, is direct in that it is unmediated by any conscious conception of how I ought to feel or behave under these circumstances. And this affective motivational state in turn causes me to perform a purposeful action, namely, to scatter dollar bills in the street. But this action is equally unmediated by any desire or "pro-attitude" toward scattering dollar bills in the street, for I would feel no frustration or regret if I were prevented from doing so.[14] My motive for doing so is that I am feeling expansive. And I was caused to feel expansive by having just got a raise.

Thus a backward-looking motive (my feeling of expansiveness) can cause purposeful action (scattering dollar bills in the street) without the intervention of a forward-looking motive. Some other examples: free-floating anxiety, consequent on my perceived social incompetence, causes me to roll my napkin into little balls at dinner; irritation at my government's obtuseness causes me to bang the plates and cutlery while setting the table; fear, consequent on my awareness that I could be hauled into court by the Internal Revenue Service for income-tax evasion, causes me to pay my taxes. Each of these is an

[13] I defend this claim at somewhat greater length in "Two Conceptions of the Self," *Philosophical Studies*, XLVIII, 2 (September 1985): 173–197; reprinted in *The Philosopher's Annual*, VIII (1985).

[14] And again: even if I did feel frustration or regret at being thus prevented, it would still be a moot question whether it was my pro-attitude toward scattering dollar bills in the street, or my feeling expansive, that caused me to scatter dollar bills in the street.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

example of purposeful action motivated by backward-looking affectively motivating states.

Similarly, feelings of respect for the moral imperative to aid imperiled loved ones first, consequent on my awareness of Jeff as an imperiled loved one, causes me to save Jeff first. An intentional object, i.e., a loved one's peril and my primary obligation to aid him, causes a backward-looking affectively motivating state, i.e., respect, which in turn causes a purposeful action, i.e., my saving Jeff first. Now I feel respect for imperatives thus derived from my moral theory, because I feel both the force of logic, and the immediacy of the application of this theory to our situation: My moral theory governs my understanding of the events I perceive—i.e., that Jeff is imperiled and that I must save him right away—and it motivates my responses to them—i.e., my direct and unambivalent attempt to save him. But it would have neither of these features if it furnished no guidance for the treatment of loved ones, nor for rendering aid to the imperiled. And of course no one would be tempted to take seriously a moral theory as impoverished as this. Only a theory capable of guiding and making sense of moral experience *in practice* can elicit our respect.

I could not, however, identify Jeff as an imperiled loved one relative to my respected moral theory, were it not for my prior, unmediated love and concern for him. Moral-alienation critics tend to speak as though to have an overriding personal investment in an impartial moral theory is not only to suppose that moral principles apply to all human agents (true), but to be motivated primarily by concern to conform to this theory to enter into personal relationships in the first place[15] (false). To be sure, if I am in fact a moral person, then moral principles apply to my personal relations, and my behavior toward others either exemplifies or violates these principles, regardless of any changes in my attitude toward either: All is fair neither in love nor in war. If, further, I identify myself as a moral person, then the principles derived from my moral theory not only

[15] Bernard Williams, for example, claims that, for the Kantian, "personal relations at least presuppose moral relations. . . . [T]hey are applications to this case of relations which the lover, *qua* moral person, more generally enters into" ("Person, Character and Morality," *op. cit.*, p. 16). Similarly, Michael Stocker argues as follows: "Suppose you embody this utilitarian reason as your motive in your actions and thoughts toward someone. Whatever your relation to that person, it is necessarily not love (nor is it friendship, affection, fellow feeling, or community). The person you supposedly love engages your thought and action not for him/herself, but rather as a source of pleasure" ("The Schizophrenia of Modern Ethical Theories," *op. cit.*, p. 458).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

apply to my personal relations, but also guide them. But that moral imperatives *guide* my personal relations cannot imply that moral imperatives are *presupposed by* my personal relations. For if we could have no personal relations without presupposing moral relations, there would be no examples for the imperatives that define moral relations to apply to: if I did not already love Jeff (and recognize his peril), I could not apply the moral imperative to aid imperiled loved ones first to our situation. And if I bore no such personal relation to anyone, obviously this imperative could have no application at all.

Hence my respect for this imperative need not blind me to Jeff's uniqueness, nor pre-empt my affection for him, any more than my impartial belief that smoking is unhealthy blinds me to the temptation of the cigarette before me, or pre-empts the craving to which I am in danger of succumbing. It is an interesting view of moral obligations that regards them as stifling or distorting our personal relationships; as though the obligation to treat a loved one with special care somehow took all the fun out of it.

Moreover, it is precisely my respect for this moral imperative that obviates any doubt or ambivalence that might otherwise cause me to hesitate in deciding whom to save first. If I didn't respect my special moral obligation to loved ones, my disposition to save Jeff first might be overriding, but it would not be unqualified by ambivalence about where my moral obligation lay. Without my recognition of Jeff as a loved one, my disposition to save him first might not be qualified by qualms about my moral duty, but it might not be overriding either. Being motivated to save Jeff first by this moral imperative, then, as much presupposes an unmediated personal relationship to Jeff as it does respect for my moral theory.

Some might maintain that it is precisely the potential for ambivalence, or for a conflict between love and duty, that shows the fundamental defect of impartial moral imperatives. That they might prescribe one course of action, and my natural inclinations another, reinforces the alienation that, I claim, is a straw man. But the problem is then not local to impartial moral prescriptions, but instead common to any morality—indeed, to any prescriptions of any kind that happen to diverge from what I am naturally inclined to do.[16] If we think of a morality as, roughly, a way in which our actions and emotions are or should be regulated by the legitimate requirements

[16] Marcia Baron touches on this point in "The Alleged Repugnance of Acting from Duty," this JOURNAL, LXXXI, 4 (April 1984): 197–220, p. 213.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of others, then the objection is, in fact, an objection to heeding those requirements at the expense of one's personal inclinations, and a complaint that one is not invariably encouraged to indulge them. But I have already suggested that such complaints ultimately support pathological narcissism.

A motivationally effective moral imperative, then, ordinarily presupposes rather than precludes unmediated feelings of concern or affection. So to be *motivated* to save Jeff first by respect for a moral imperative does not imply that my *purpose* in acting is to obey that imperative to the detriment of my overriding concern for Jeff, any more than being motivated by fear of the IRS to pay my taxes implies that my purpose in acting is to obey the IRS to the detriment of my overriding concern to pay my taxes. In both cases, my complex response to a perceived intentional object (the specter of the IRS, a loved one's peril) includes a backward-looking affectively motivating state (fear of the IRS, respect for the moral law) that motivates purposeful action (paying my taxes, saving Jeff first).

Let us then think of a *morally integrated agent* as, very roughly, one whose dispositions, prescribed by her moral theory, are sufficiently deeply instilled, preferably in the normal process of socialization, as to be motivationally effective most of the time.[17] This means, first, that her actions usually, although not invariably, conform to its prescriptions. Second, it means that her actions are invariably guided by these prescriptions in the following ways:

First, she naturally develops relationships with others that elicit mutual trust, affection, respect, etc., or their opposites, and interprets these relationships, actions, emotions, and individuals with the help of the impartial vocabulary and concepts her moral theory supplies. Thus she views people's actions, her own included, as right or wrong, well intended or malificent, honorable or shameful, and so on; and people themselves, herself included, as accordingly judicious or partial, benevolent or malevolent, virtuous or corrupt, generous or spiteful, good or bad, and so on. That is, she recognizes the terms and prescriptions of her moral theory to apply to her experience.

Second, these morally theory-laden judgments reinforce some affectively or perceptually motivating states at the expense of others and some behavioral dispositions at the expense of others. Thus, for example, her judgment that she is selfish makes her feel ashamed, and so motivates her to behave unselfishly; her judgment that others

[17] This very rough sketch is amplified in greater detail in my *Rationality and the Structure of the Self*, XII, *op cit.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

are beneficent disposes her to reciprocate. That is, her morally theory-laden experiences reinforce or undermine her moral training.

On this skeletal account, it would be misleading to deny that an agent has a conscious commitment to her moral theory; for its concepts and prescriptions saturate her interpretation of morally appropriate behavior, of her own emotions and actions, and of herself and other people. She thinks of them as, for example, friends, responsible agents, rational beings, loved ones, etc. But it would be similarly misleading to complain that her moral theory alienates her from the objects of her moral concern. For it is only with the aid of her moral theory that she is able to recognize situations as being those in which moral concern is appropriate. Without her moral theory, she would lack the concept of a person as good, valuable, a friend, or deserving of aid or respect. Without these concepts, it is unclear what would cause her to feel respect, compassion, or sympathy for him. Then it is equally unclear what would motivate her moral behavior. But of course to be consciously committed to a moral theory does not imply that one must be constantly preoccupied with it. On this account, our moral behavior is motivated by our theory-laden perception of moral situations as moral. Our moral behavior successfully conforms to the prescriptions of our moral theory to the extent that this theory is sufficiently fine-tuned, that its prescriptions are coordinated with our normative dispositions, and that these dispositions have been fully internalized.

All of this is not to deny that moral alienation is a problem for some agents. But I have tried to show that whether our sensitivity and our vision of others is obscured or clear, and our social relations alienated or integrated, is largely independent of whether our view of the world is theoretical or affective. So although moral alienation may well be a problem for some agents, it is not a *moral* problem.

ADRIAN M. S. PIPER

Georgetown University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Memo to Our Readers, concerning Copyright:

The editors and trustees of *The Journal of Philosophy* wish to draw to the attention of our readers the notice of permission for personal or educational copying which appears at the bottom of the masthead page of each issue of this JOURNAL.

In the last few years we have received many requests for permission to photocopy or mimeograph articles for classroom use; we have also received many requests from libraries for blanket permission to photocopy. And we are aware of course that a great deal of such photocopying has been done without permission. We are grateful to those of you who have asked for our permission.

The copyright law of 1976 grants certain new rights to libraries and educators who need to photocopy and certain new protections to copyright owners. In general, photocopies in excess of "fair use" copying cannot be made without authorization by the copyright owner. What constitutes "fair use" is not yet entirely clear, but some single-copying for personal use is permitted, and multiple-copying for classroom use is clearly limited: the article copied must be *short* (less than 2,500 words) or only an excerpt copied; the copying must be *spontaneous* (at the instance of the individual instructor, and needed so promptly as not to allow time to obtain permission); and *each copy must include a notice of copyright.*

The Association of American Publishers has been instrumental in setting up an organization, the Copyright Clearance Center, designed to make it as simple as possible for organizations to pay the fees required for copying *beyond fair use.* The *Journal* has joined the Center and thereby authorizes such copying under the terms established by the Center. The per-copy price for personal or educational copying beyond fair use appears on the first page of each article in the *Journal.* That price is 10¢/page, with a minimum of 50¢, unless some special arrangement is requested by the author. The *Journal* continues to give permission for multiple-copying for classroom use, but at 5¢/copy/page, where an instructor applies directly to us for our permission. We turn over to our authors all substantial sums ($15 or more) obtained from their work.

We wish also to point out that, in most cases, the cost of photocopying a *Journal* article is higher than the price of a *Journal* issue, even with our new monthly format ($2.00 for a current issue, $3.00 for a back issue of 64 pages). *Journal* issues are available from this office, singly as well as in volumes, back to 1904; prepaid orders are filled by return mail.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

The American Society for Aesthetics announces a call for papers for its Forty-fifth Annual Meeting, to be held October 28–31, 1987, at the University of Missouri in Kansas City, Missouri. The deadline for submissions is March 1987. Further information may be obtained from: Allen Carlson, Department of Philosophy, University of Alberta, Edmonton, Alberta T6G 2E5, Canada.

The Fourteenth Annual Saint Louis Conference on Manuscript Studies will be held at St. Louis University on October 16 and 17, 1987. An invitation has been extended for papers dealing with one of the four following aspects of the manuscript: codicology, illumination, paleography, and texts. Additional information may be obtained from the Conference Committee, *Manuscripta,* Vatican Film Library, 3650 Lindell Blvd., St. Louis, Mo. 63108.

The Association for Philosophy of Education is pleased to announce a call for papers for its first meeting, to be held in conjunction with the Eastern Division Meeting of the American Philosophical Association in December, 1987. Papers should present philosophical analyses pertaining to some aspect of education. The deadline for submission of papers is March 31, 1987. Submissions may be sent to Sophie Haroutunian-Gordon, Department of Education, University of Chicago, 5835 South Kimbark Avenue, Chicago, Ill. 60637, or to Harvey Siegel, Department of Philosophy, University of Miami, Coral Gables, Fla. 33124.

Oxford University Press announces a new series of Monographs in the History and Philosophy of Biology. Under the direction of Richard M. Burian (Virginia Polytechnic and State University), Richard W. Burkhardt (Illinois/Urbana), Richard C. Lewontin (Harvard), and John Maynard Smith (Sussex), this series will be a central resource for those interested in the philosophical and historical foundations of biology and with current controversy in the field. Inquiries and manuscript submissions may be directed to: William Curtis, Editor, Oxford University Press, 200 Madison Avenue, New York City, 10016.

The Society for Philosophy and Public Affairs, New York group, is pleased to announce its Spring Colloquium schedule. Feb. 19: Sally Kagan (Illinois/Chicago Circle) "Libertarianism and Freedom"; March 18: Barbara Herman (Princeton) "Violence and Self-defense in Kantian Ethics"; April 18: Arthur Kuflik (Vermont) "Intellectual Property," with Stefan Bauer Mengelberg (SUNY/Binghamton) commenting; May 12: William Ruddick (NYU) "The Limits of Professionalism." All colloquia will be held at the Loeb Student Center, Room 409, 566 LaGuardia Place, at 7:15 P.M. Further information may be obtained from: Gertrude Ezorsky, Philosophy Department, Brooklyn College, N.Y. 11210.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The Social Philosophy and Policy Center of Bowling Green State University is pleased to announce the appointment of R. G. Frey as Senior Research Fellow.

The Center also announces a conference devoted to David Gauthier's book, *Morals by Agreement,* to be held April 3/4, 1987, at Bowling Green State University. Speakers will include Russell Hardin (Chicago), Kurt Baier (Pittsburgh), Gilbert Harman (Princeton), James Fishkin (Philosophy and Government/Texas at Austin), Laurence Thomas (Oberlin), Christopher Morris (Bowling Green), and James Buchanan (Economics/George Mason). David Gauthier (Pittsburgh) will reply to the papers presented. More information may be obtained from: Kory Tilgner, Social Philosophy and Policy Center, Bowling Green State University, Bowling Green, Ohio 43403.

The Humanities Collegium of the Claremont Graduate School invites applications for a two-year post-doctoral fellowship in the humanities, beginning in the Autumn semester, 1987. Fellows will have the opportunity to pursue their research and to teach a course related to that research; they will also assist the Director of the Collegium in conducting the Commission on Humanities and Public Affairs, a group of citizens from business and the professions who meet once a month with faculty and students to discuss major social issues. Next year's theme will be chosen to commemorate the bicentennial of the Constitution. Inquiries may be addressed to: A. R. Louch, Director, Humanities Collegium, The Claremont Graduate School, Claremont, Calif. 91711.

The Department of Philosophy of the University of Kansas is pleased to announce that Jerzy Pelc, of the Faculty of Philosophy and Sociology at the University of Warsaw, will be the Rose Morgan Visiting Distinguished Professor during the Spring semester, 1987. Professor Pelc is President of the International Institute of Philosophy and of the International Association of Semiotic Studies.

The Department of Philosophy of the University of Miami will hold a conference in Coral Gables on The Background of Contemporary Philosophical Logic, to be held March 13/14, 1987. Participants will be: Arthur Burks (Ann Arbor): "Peirce's and Marquand's Ideas for Electrical Computers"; Rudolf Haller (Graz): "Incompleteness and Fictionality in Meinong's Object Theory"; Risto Hilpinen (Miami and Turku): "Peirce, Goodman, and the Aesthetic Sign"; Jaakko Hintikka (Tallahassee): "The Universality of Language and the Model-theoretic Point of View"; Ruth Barcan Marcus (Yale): "C. I. Lewis's Modal Logic"; Charles G. Morgan (Victoria): "On Peirce's Philosophical Logic"; Seppo Sajama (Turku): Meinong on the Foundations of Deontic Logic"; and Thomas Short (Gambier): "Hypostatic Abstraction in Empirical Science." More information may be obtained from: Howard Pospesel, Department of Philosophy, University of Miami, P.O. Box 248054, Coral Gables, Fla. 33124.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The Center for Philosophy of Science at the University of Pittsburgh announces a call for papers for its Sixth Annual Conference on Philosophy of Science, to be held December 11/12, 1987. This year's conference, entitled "Aesthetics's Factors in Natural Science", will examine the place of aesthetic parameters in scientific explanation: simplicity, uniformity, symmetry, economy, and elegance, for example. Deadline for submissions is May 15, 1987. Submissions and inquiries may be addressed to: Nicholas Rescher, Director, Center for Philosophy of Science, 817 Cathedral of Learning, University of Pittsburgh, Pittsburgh, Pa. 15260.

The Department of Philosophy of Virginia Polytechnic Institute is pleased to announce a conference on Individual Autonomy, to be held May 15–17, 1987. Participants include Gerald Dworkin (Illinois/Chicago), David A. J. Richards (NYU/Law), Russell Hardin (Chicago), James Rachels (Alabama/Birmingham), William Ruddick (NYU), Susan Wolf (Johns Hopkins), Thomas Hill (North Carolina). Commentators are Eleonore Stump, William Williams, James Klagge, and John Christman, all of Virginia Polytechnic Institute. More information may be obtained from John Christman, Department of Philosophy, Virginia Polytechnic Institute, Blacksburg, Va. 24061.

The Department of Philosophy, University of Western Ontario, will hold a conference on contemporary contractarian moral philosophy, April 24–26, 1987. Participants include: David Braybrooke (Dalhousie), David Copp (Illinois/Chicago and Simon Fraser), Peter Danielson (York), David Gauthier (Pittsburgh), Allan Gibbard (Michigan/Ann Arbor), Jean Hampton (California/Los Angeles), Russell Hardin (Chicago), Virginia Held (CUNY), Jan Narveson (Waterloo), John Roemer (California/Davis), Geoffrey Sayre-McCord (North Carolina/Chapel Hill), and Holly Smith (Arizona). More information may be obtained from Peter Vallentyne, Department of Philosophy, University of Western Ontario, London, Ontario N6A 3K7, Canada.

The editors wish to remind our correspondents that notices of scheduled meetings and symposia are of value to our readers only if they appear early enough to make attendance possible. Copy for Notes and News must go to the printer at least six weeks before publication date, and the Notes and News column must be planned and processed in this office. Accordingly, an announcement of a meeting to be held, say, in April, should be scheduled for our February issue (or sooner) and should be received in this office no later than December 1, that is, at least 4 months in advance. The editors appreciate notices that come to us ready for the printer, requiring little editorial work and no retyping, in JOURNAL style (see these columns), with no imperatives, no prices, no telephone numbers, no headings, no unnecessary capitals; and reasonably terse, since space is limited.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIV, NO. 3, MARCH 1987

INDETERMINACY, EMPIRICISM, AND THE FIRST PERSON*

THE aim of this article is to assess the significance of W. V. Quine's indeterminacy thesis. If Quine is right, the thesis has vast ramifications for the philosophy of language and mind; if he is wrong, we ought to be able to say exactly how and why.

I

Let us begin by stating the behaviorist assumptions from which Quine originally proceeds. For the sake of developing an empirical theory of meaning, he confines his analysis to correlations between external stimuli and dispositions to verbal behavior. In thus limiting the analysis, he does not claim to capture all the intuitions we have about the pretheoretical notion, but rather the "objective reality"[1] that is left over if we strip away the confusions and incoherencies in the pretheoretical "meaning." The point of the "behavioristic ersatz" is to give us a scientific, empirical account of the objective reality of meaning. On this view, the objective reality is simply a matter of being disposed to produce utterances in response to external stimuli. The stimuli are defined entirely in terms of patterns of stimulations of the nerve endings, and the responses entirely in terms of sounds and sound patterns that the speaker is disposed to emit. But we are not supposed to think that between the stimulus and the verbal response there are any mental entities. We are not supposed to think that there is any consciousness, intentionality, thoughts, or any internal "meanings" connecting the stimuli to the noises. There

* I am indebted to a large number of people for comments and criticism of earlier drafts of this paper. I especially want to thank Noam Chomsky, Dagfinn Føllesdal, Ernest Lepore, Brian McLaughlin, George Myro, Dagmar Searle, and Bruce Vermazen.

[1] *Word and Object* (Cambridge, Mass.: MIT Press; New York: Wiley, 1960), p. 39.

0022-362X/87/8403/0123$02.40 © 1987 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is just the pattern of stimulus and the pattern of learned response. There will, of course, be neurophysiological mechanisms mediating the input and the output, but the details of their structure do not matter to a theory of meaning, since any mechanism whatever that systematically associated stimulus and response would do the job as well. For example, any computer or piece of machinery that could emit the right sounds in response to the right stimuli would have "mastered" a language as well as any other speaker, because that is all there is to the mastery of a language. Quine, I take it, does not deny the existence of inner mental states and processes; he just thinks they are useless and irrelevant to developing an empirical theory of language.

Such a view is linguistic behaviorism with a vengeance. It has often been criticized and, in my view, often refuted, for example, by Noam Chomsky in his review of B. F. Skinner.[2] On one construal, my Chinese room argument can also be interpreted as a refutation.[3] One way to refute this version of extreme linguistic behaviorism (let us call it "behaviorism" for short) would be to offer a reductio ad absurdum of its basic premises; and, indeed, it seems to me that Quine has offered us one such famous reductio (*op. cit.*, ch. 2). If behaviorism were true, then certain distinctions known independently to be valid would be lost. For example, we all know that, when a speaker utters an expression, there is a distinction between his meaning rabbit and his meaning rabbit stage or undetached rabbit part. But, if we actually applied the assumptions of behaviorism to interpreting the language of an alien tribe, we would find there was no way of making these distinctions as plain facts of the matter about the language used by the native speakers. Suppose, for example, the natives shouted "Gavagai!" whenever a rabbit ran past, and suppose we tried to translate this into our English as "There's a rabbit!" or simply, "Rabbit!". The stimulus—which, remember, is defined entirely in terms of stimulations of nerve endings—is equally appropriate for translating "Gavagai!" as "There's a stage in the life history of a rabbit!" or "There's an undetached part of a rabbit!". The

[2] "Review of B. F. Skinner's *Verbal Behavior*," in Jerry Fodor and Jerrold Katz, eds., *The Structure of Language* (Englewood Cliffs, N.J.: Prentice-Hall, 1964), pp. 547–578.

[3] In the Chinese room argument, the man in the room follows a computer program that makes his verbal behavior indistinguishable from that of a Chinese speaker, but he still does not understand Chinese. He satisfies the behavioral criterion for understanding without actually understanding. Thus, the refutation of strong AI is a fortiori a refutation of behaviorism. [See my "Minds, Brains, and Programs," *Behavioral and Brain Sciences*, III (1980): 417–457; and *Minds, Brains, and Science* (Cambridge, Mass.: Harvard, 1984).]

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

same pattern of stimulation of the photoreceptor cells does duty for all three translations. So, if all there were to meaning were patterns of stimulus and response, then it would be impossible to discriminate meanings, which are in fact discriminable. That is the reductio ad absurdum.

It is crucial to this argument to see that, even if we got more patterns of stimulus and response for our tribe, that still would not enable us to make the discriminations we need to make. Suppose we learned their expression for 'is the same as' and tried to use it to enable us to tell whether they meant rabbit or rabbit stage or undetached rabbit part. We could get the rabbit to run past again, and if they said "Same gavagai," we would have at least pretty good evidence that they did not mean, for example, rabbit stage by 'gavagai'. But this would be no help to us at all, because exactly the same sorts of doubt that we had about 'gavagai' in the first place would now apply to the expression for 'is the same as'. As far as matching stimuli and responses is concerned, we could equally well translate it as 'is a part of' or 'belongs with'. The conclusion we are forced to is this: assuming linguistic behaviorism, there will be endlessly different and inconsistent translations, all of which can be made consistent with all actual and possible evidence concerning the totality of the speech dispositions of the native speakers. As far as the behavioral evidence is concerned, there is nothing to choose between one translation and another even though the two are inconsistent.[4]

On Quine's view, the unit of analysis for empirically testing translations is not words or individual expressions but whole sentences. The only direct empirical checks we have on translations are for those sentences which are associated directly with stimulus conditions, the "observation sentences." On this view, 'Gavagai!', 'Rabbit!', 'Rabbit stage!', 'Undetached rabbit part!' all have the same determinate stimulus meaning; they have "stimulus synonymy," since the same stimulus conditions would prompt assent to or dissent from them. The indeterminacy arises when we attempt to form "analytical hypotheses" that state the meanings of particular words or other elements of the sentence. The indeterminacy that attaches to the elements of observation sentences is at least constrained by the stimulus conditions that prompt assent to or dissent from those

[4] In what sense exactly can two translations be inconsistent? We cannot simply say that they have different *meanings*, for that would seem to imply the existence of determinate meanings. Rather, we must say that they are inconsistent in the sense that one system of translation will accept translations that the other system would reject [Quine, "Reply to Harman," *Synthese*, XIX, 1/2 (December 1968): 267–269; also, *Word and Object*, pp. 73/4.]

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sentences. The determinate stimulus meaning that attaches to observation sentences should at least seem puzzling to us, however, since sentences that have the same stimulus meaning do not in any ordinary sense of 'meaning' have the same meaning. By any reasonable standard of objective reality, it is a matter of objective reality that "There's a rabbit" and "There's an undetached rabbit part" just do not mean the same things. The significance of this point for the over-all theory will emerge later.

Now, why exactly is Quine's argument a reductio ad absurdum of extreme linguistic behaviorism? There are two positions which are inconsistent:

(1) The thesis of behaviorism: The objective reality of meaning consists entirely of correlations between external stimuli and dispositions to verbal behavior.[5]

(2) In a given case of speech behavior, there can be a plain fact of the matter about whether a native speaker meant, e.g., rabbit, as opposed to rabbit stage, or undetached rabbit part, by the utterance of an expression.

If alternative and inconsistent translation schemes can all be made consistent with the same patterns of stimulus and response, then there cannot be any fact of the matter about which is right, because, according to (1), there isn't anything else to be right about. But this is inconsistent with (2); so if we accept (2), (1) must be false.

I think it is clear which of (1) or (2) we have to give up. Quine has simply refuted extreme linguistic behaviorism. But why am I so confident about that? Why not give up (2)? The answer is the obvious one: if behaviorism were correct, it would have to be correct for us as speakers of English as well as for speakers of Gavagai-talk. And we know from our own case that we do mean by 'rabbit' something different from 'rabbit stage' or 'undetached rabbit part'. If my English-speaking neighbor, having read Quine, decides that he can't tell whether by 'rabbit' I mean rabbit, undetached rabbit part, or rabbit stage, then so much the worse for him. When I saw a rabbit recently, as I did in fact, and I called it a rabbit, I meant rabbit. In all discussions in the philosophy of language and the philosophy of mind, it is absolutely essential at some point to remind oneself of the first-person case. No one, for example, can convince us by argument, however ingenious, that pains do not exist if in fact we have them, and similar considerations apply to Quine's example. If somebody has a theory according to which there isn't any difference between my

[5] Sometimes Quine talks about behavior *simpliciter*, sometimes about *dispositions* to behavior. I think the notion of dispositions to behavior is the one he prefers.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

meaning rabbit and my meaning rabbit part, then I know that his theory is simply mistaken; and the only interest his theory can have for me is in trying to discover where he went wrong. I want to emphasize this point, since it is often regarded as somehow against the rules in these discussions to raise the first-person case.

In a different philosophical environment from the one we live in, this might well be the end of the discussion. Linguistic behaviorism was tried and refuted by Quine using reductio ad absurdum arguments. But, interestingly, he does not regard it as having been refuted. He wants to hold behaviorism, together with the conclusion that, where analytical hypotheses about meaning are concerned, there simply are no facts of the matter, together with a revised version of (2), the thesis that we can in fact make valid distinctions between different translations. And some authors, such as Donald Davidson[6] and John Wallace,[7] who reject behaviorism, nonetheless accept a version of the indeterminacy thesis. Davidson, in fact, considers and rejects my appeal to the first-person case. Why does the thesis of the indeterminacy of translation continue to be accepted? And what larger issues are raised by the dispute? I now turn to these questions.

II

We need to consider three theses:

(A) The indeterminacy of translation
(B) The inscrutability of reference
(C) The relativity of ontology

In this section, I will first explain the relations between (A) and (B), and then try to say more about the character of the thesis Quine is advancing. In the next section, I will try to show that (C) is best construed as an unsuccessful maneuver to rescue the theory from the apparently absurd consequences of (A) and (B).

The thesis of the indeterminacy of translation is that, where questions of translation and, therefore, of meaning are concerned, there is no such thing as getting it right or wrong. This is not because of an epistemic gulf between evidence and conclusion, but because there is no fact of the matter to be right or wrong about.

From (A), so stated, (B) follows immediately. For if there is no fact of the matter about whether or not a speaker *meant* rabbit as op-

[6] "The Inscrutability of Reference," *Southwestern Journal of Philosophy*, X (1979): 7–19, reprinted in *Inquiries into Truth and Interpretation* (New York: Oxford, 1984), pp. 227–241; page references are to this version.

[7] "Only in the Context of a Sentence Do Words Have Any Meaning," *Midwest Studies in Philosophy*, II: *Studies in the Philosophy of Language* (1977).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

posed to rabbit stage, then equally, there is no fact of the matter about whether or not he is *referring* to a rabbit or a rabbit stage. In Fregean terminology, indeterminacy of sense entails inscrutability of reference.

Now, if we were to construe (A) as just the claim that there are no psychological facts of the matter about meanings in addition to facts about correlations of stimulus and response, then it would seem puzzling that we didn't derive that conclusion immediately from extreme linguistic behaviorism. It would seem puzzling that there is so much heavy going about 'gavagai', etc. But thesis (A) is stronger than just the thesis of behaviorism; that is, it is stronger than the claim that there isn't any meaning in addition to correlations of stimulus and response. It says further that there are an indefinite number of equally valid but inconsistent ways of correlating stimulus and verbal response in the vocabulary of an alien language with that of our language. The thesis that there are no objectively real meanings in addition to dispositions to verbal behavior was already assumed at the beginning of the discussion. Quine rejected any appeal to meanings, in any psychological sense, from the start. That was never at issue. What was at issue was the possibility of empirically motivated correct translations from one language to another, *given behaviorism;* the issue was whether or not there is an empirically motivated notion of sameness of meaning left over after we have adopted extreme linguistic behaviorism.

We will see the importance of this consideration when we see why several criticisms that are made of Quine miss the mark. Chomsky, for example, has repeatedly claimed that Quine's thesis of indeterminacy is simply the familiar underdetermination of hypothesis by empirical evidence.[8] Because any empirical hypothesis makes a claim that goes beyond the evidence, there will always be inconsistent hypotheses that are consistent with any actual or possible evidence. But underdetermination, so construed, does not entail that there is "no fact of the matter." Now Quine's response to Chomsky's objection seems at first sight puzzling. He grants that indeterminacy is underdetermination, but claims that it is underdetermination at one remove and, therefore, that there is no fact of the matter. He claims that, even if we have established all the facts about physics, semantics is still indeterminate. He writes:

> Then when I say there is no fact of the matter, as regards, say, the two rival manuals of translation, what I mean is that both manuals are com-

[8] Cf., for example, his "Quine's Empirical Assumptions," *Synthese,* XIX, 1/2 (December 1968): 53–68.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

patible with all the same distributions of states and relations over elementary particles. In a word, they are physically equivalent.[9]

But this answer seems inadequate to Chomsky and at one time seemed inadequate to me, because underdetermination at one remove is still just underdetermination. It wouldn't be sufficient to show that there is no fact of the matter. The objection to Quine that Chomsky makes (and that I used to make) is simply this: for any given higher-level "emergent" or "supervenient" property, there will be (at least) two levels of underdetermination. There will be a level of the underdetermination of the underlying physical theory, but there will also be a theory at the higher level, for example, at the level of psychology; and information at the level of microphysics is, by itself, not sufficient to determine the level of psychology. As Chomsky once put it, if you fix the physics, the psychology is still open; but equally, if you fix the psychology, the physics is still open. For example, the theory of all the dispositions of physical particles that go to make up my body, by itself, would leave open the question of whether or not I am in pain. The thesis that I am in pain is underdetermined at one remove. Now why is it supposed to be any different with meaning? Of course, there are two levels of underdetermination, but in both cases there are facts of the matter—in one case, facts of psychology, and in the other case, facts of physics. I now believe that this answer misses Quine's point altogether because it fails to see that he is assuming from the start that there is no psychologically real level of meaning beyond simple physical dispositions to respond to verbal stimuli. To repeat, Quine assumes from the very start the nonexistence of (objectively real) meanings in any psychological sense. If you assume that they are so much as possible, his argument fails. But now it begins to look as though the real issue is not about indeterminacy at all; it is about extreme linguistic behaviorism.

Many philosophers assume that Quine's discussion is sufficient to refute any sort of mentalistic or intentionalistic theory of meaning. But what our discussion of Chomsky's objections suggests is that this misconstrues the nature of the discussion altogether. It is only *assuming* the nonexistence of intentionalistic meanings that the argument for indeterminacy succeeds at all. Once that assumption is abandoned, that is, once we stop begging the question against mentalism, it seems to me that Chomsky's objection is completely valid. Where meanings psychologically construed are concerned, there is the familiar underdetermination of hypothesis by evidence, and that underdetermination is in addition to the underdetermination at the

[9] *Theories and Things* (Cambridge, Mass.: Harvard, 1981), p. 23.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

level of physical particles or brute physical behavior. So what? These are familiar points about any psychological theory. There is nothing special about meaning and nothing to show that where meaning is concerned there is no fact of the matter.

To deepen our understanding of these points, we must now turn to the thesis of the relativity of ontology.

III

Quine recognizes that the proofs of the indeterminacy of translation and of the inscrutability of reference seem to be leading to absurd consequences. He writes:

> We seem to be maneuvering ourselves into the absurd position that there is no difference on any terms, interlinguistic or intralinguistic, objective or subjective, between referring to rabbits and referring to rabbit parts or stages; or between referring to formulas and referring to their Gödel numbers. Surely this is absurd, for it would imply that there is no difference between the rabbit and each of its parts or stages, and no difference between a formula and its Gödel number. Reference would seem now to become nonsense not just in radical translation but at home.[10]

The indeterminacy thesis seems to have the absurd consequence that indeterminacy and inscrutability apply to the first-person case, to oneself: "If it is to make sense to say even of oneself that one is referring to rabbits and formulas and not to rabbit stages and Gödel numbers, then it should make sense equally to say it of someone else" (*ibid.*, 47).

Quine recognizes something that many of his critics have missed, and that is the real absurdity of the indeterminacy argument once you follow out its logical consequences: followed to its conclusion, the argument has nothing essentially to do with translating from one language to another or even understanding another speaker of one's own language. If the argument is valid, then it must have the result that there isn't any difference *for me* between *meaning* rabbit or rabbit stage, and that has the further result that there isn't any difference for me between *referring to* a rabbit and referring to a rabbit stage, and there isn't any difference for me between something's *being* a rabbit and its *being* a rabbit stage. And all of this is a consequence of the behaviorist assumption that there isn't any meaning beyond behaviorist meaning. Once we concede that as far as behaviorist "stimulus meaning" is concerned, 'There's a rabbit' and 'There's a rabbit stage' are "stimulus synonymous," then the rest

[10] *Ontological Relativity and Other Essays* (New York: Columbia, 1969), pp. 47/8.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

follows, because on the behaviorist hypothesis there isn't any other kind of objectively real meaning or synonymy. I think, with Quine, that these consequences are absurd on their face, but if there is any doubt about their absurdity, recall that the whole argument about 'Gavagai' was understood by me (or you) only because we know the difference for our own case between meaning rabbit, rabbit stage, rabbit part, etc.

I said in the last section that the thesis of indeterminacy is the thesis that there cannot be empirically well-motivated translations of the words of one language into those of another, given behaviorism. But if this thesis is correct, then there cannot even be "correct" translations from a language into itself. By observing my idiolect of English, I can't tell whether by 'rabbit' I mean rabbit stage, rabbit part, or whatnot. Quine need not have considered Gavagai speakers. He could have simply observed in his own case that there was no "empirical" difference between his meaning one thing or the other and, therefore, that there was no real difference at all. And that result, as he correctly sees, is absurd. *If the indeterminacy thesis were really true, we would not even be able to understand its formulation; for when we were told there was no "fact of the matter" about the correctness of the translation between rabbit and rabbit stage, we would not have been able to hear any (objectively real) difference between the two English expressions to start with.*

Here is Quine's picture: I am a machine capable of receiving "nerve hits" and capable of emitting sounds. I am disposed to emit certain sounds in response to certain nerve hits; and, objectively speaking, that is all there is to meaning. Now the stimulus meaning of "There's a rabbit stage" is the same as that of "There's a rabbit," since the sounds are caused by the same nerve hits. It isn't just that Quine has a technical notion of "stimulus meaning" which he wants to add to our common-sense notion of meaning. No, he thinks that, as far as objective reality is concerned, stimulus meaning is all the meaning there is. And it is his notion of stimulus meaning which generates the absurdity.

The resolution of this "quandary," according to Quine, lies in perceiving the *relativity* of reference and ontology. "Reference *is* nonsense except relative to a coordinate system" (*ibid.*, 47), and the coordinate system is provided by a background language. The question for me of whether I am referring to a rabbit by 'rabbit' is answered by simply taking the English background language for granted, by "acquiescing in our mother tongue and taking its words at face value" (49). Just as in physics it makes sense to speak of the position and velocity of an object only relative to a coordinate sys-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tem, so analogously it makes sense to talk of the reference of an expression only relative to some background language. Indeed, where translation from another language is concerned, reference is doubly relative: relative first to the selection of a background language into which to translate the target language, and relative second to the arbitrary selection of a translation manual for translating words of the target into the background.

Now, does this answer remove the apparent absurdity? I do not see how it does; indeed I shall argue that it simply repeats the problem without solving it.

I believe that with the thesis of relativity we have reached the crux of the indeterminacy argument. For this issue we can forget all about 'gavagai' and radical translation; they were merely picturesque illustrations of the consequences of behaviorism. The crucial thesis can be exemplified as follows:

> There is no empirical difference between the claim that I meant rabbit by 'rabbit' and the claim that I meant, e.g., rabbit stage by 'rabbit'.

This is a consequence of the original thesis of *Word and Object,* and it is now admitted to be absurd. So to get out of the absurdity we substitute a revised relativity thesis:

> Relative to one arbitrarily selected translation scheme we can truly say that I meant rabbit, relative to another scheme, equally arbitrary, that I meant, e.g., rabbit stage, *and there is no empirical difference between the two schemes.*

But the revised thesis is just as absurd as—and indeed expresses the same absurdity as—the first. And this should not surprise us, because the original absurdity arose in a discourse that already was relativized; it arose relative to my idiolect of English. The absurdity is that, if I assume my idiolect is a fixed set of dispositions to verbal behavior, then any translation of one word into itself or another of my idiolect is absolutely arbitrary and without empirical content. There is no way for me to tell whether by 'rabbit' I mean rabbit, rabbit stage, rabbit part, etc. This applies even to simple disquotation: there is no way even to justify the claim that by 'rabbit' I mean rabbit. Now, it does not meet this difficulty to say that we can fix meaning and reference by making an arbitrary selection of a translation manual. The arbitrariness of the selection of the translation manual is precisely the problem, since it is a reflection of the arbitrariness of the selection from among the original range of alternative analytical hypotheses. Quine's thesis of relativity does not remove the absurdity; it simply restates it.

When Quine advises us to acquiesce in our mother tongue and

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

take words at their face value, we have to remind ourselves that, on his account, our mother tongue consists entirely of a set of dispositions to verbal behavior in response to sensory stimuli, and, so construed, the empirical face value of 'rabbit' and that of 'rabbit stage' are indistinguishable. We really cannot have it both ways. We cannot, on the one hand, insist on a rigorous behaviorism that implies that there is no fact of the matter and then, when we get in trouble, appeal to a naive notion of a mother tongue or home language with words having a face value in excess of their empirical behavioral content. If we are serious about our behaviorism, the mother tongue is the mother of indeterminacy, and the face value is counterfeit if it suggests that there are empirical differences when in fact there are none.

But what about the analogy with physics? Will that rescue us from the absurdity? One of the peculiar features of this entire discussion is the speed with which breath-taking conclusions are drawn on the basis of a few sketchy remarks and underdescribed examples. To try to get at least a little bit clearer about what is going on, let us try to state this particular issue a little more carefully. To begin, I want to state some more of the common-sense, pre-Quinean intuitions that lead me, and to a certain extent Quine himself, to think that the theses of indeterminacy and inscrutability lead or threaten to lead to absurd results. To make it intuitively easier, let us consider the case of translation from one language to another, though it is important to remember that any difficulty we find with translation from one language to another we will also find with the case of one language alone. Let us suppose that, as I am out driving with two French friends, Henri and Pierre, a rabbit suddenly crosses in front of the car, and I declare, "There's a rabbit." Let us suppose further that Henri and Pierre do not know the meaning of the English 'rabbit', so each tries to translate it in a way that is consistent with my dispositions to verbal behavior. Henri, we may suppose, concludes that 'rabbit' means *stade de lapin*. Pierre, on the basis of the same evidence, decides it means *parti non-détachée d'un lapin*. Now according to our pre-Quinean intuitions, the problem for both Henri and Pierre is quite simple: they both got it wrong. It is just a plain fact about me that when I said "rabbit," I did not mean *stade de lapin* or *partie non-détachée d'un lapin*. Those are just bad translations. Of course, when I say that, I am making certain assumptions about the meanings of these expressions in French and, therefore, about the meanings that Henri and Pierre attach to these expressions. And these assumptions, like any other empirical assumptions, are subject to the usual underdetermination of hypotheses by evidence. Assum-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ing that I got the assumptions right, Henri and Pierre are just mistaken. But even assuming that I got my assumptions wrong, if they are wrong in a certain specific way, then Henri and Pierre are just right. That is, if, for example, Henri means by *stade de lapin* what I mean by *lapin,* then he understands me perfectly; he simply has an eccentric way of expressing this understanding. The important thing to notice is that, in either case, whether they are right about my original meaning or I am right in thinking that they are wrong, there is a plain fact of the matter to be right or wrong about.[11]

These are some of the common-sense intuitions that we need to answer. Does the analogy with the relativity of motion get us out of this quandary? Let's take the idea seriously and try it out. Suppose that in the car during our rabbit conversation Henri expresses the view that we are going 60 miles an hour, while Pierre on the other hand insists we are going only 5 miles an hour. Later it turns out that Pierre was observing a large truck we were passing and was estimating our speed relative to it, while Henri was talking about our speed relative to the road surface. Once these relativities are identified there is no longer even the appearance of paradox or disagreement. Pierre and Henri are both right. But are they analogously both right about the translation of 'rabbit' once the coordinate systems have been identified? Is it a case of moving at different semantic speeds relative to different linguistic coordinate systems? It seems to me that these absurdities are just as absurd when relativized.

On Quine's view, I am right relative to English in thinking that I meant rabbit, Pierre is right relative to French in thinking that I meant *partie non-détachée d'un lapin,* and Henri is also right relative to French in thinking that I meant *stade de lapin—even though Henri and Pierre are inconsistent with each other, and both are inconsistent with the translation I would give.* And it is not an

[11] One of the most puzzling aspects of this whole literature is the remarks people make about the ability to speak two or more languages and to translate from one to the other. Quine speaks of the "traditional equations" (*Word and Object,* p. 28) for translating from one language into another. But, except for a few odd locutions, tradition has nothing to do with it. (It is a tradition, I guess, to translate Frege's *Bedeutung* as 'reference', even though it doesn't really mean that in German.) When I translate 'butterfly' as *papillon,* for example, there is no tradition involved at all; or, if there is, I certainly know nothing of it. I translate 'butterfly' as *papillon* because that is what 'butterfly' means in French. Similarly, Michael Dummett speaks of "conventions" for translating from one language to another [see "The Significance of Quine's Indeterminacy Thesis," *Synthese,* XXVII, 3/4 (July/August 1974): 351–397]. But the point is that, if you know what the words mean, there isn't any room for further conventions. By convention, the numeral '2' stands for the number two in the Arabic notation, 'II' stands for the same number in the Roman notation. But, for these very reasons, we don't need a further convention that '2' can be translated as 'II'.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

answer to this point to maintain that the appearance of inconsistency derives from the fact that we each have different translation manuals, because the problem we are trying to deal with is that we know independently that both of their translation manuals are just plain wrong. It was the apparent wrongness of the translation manuals that we were trying to account for. To put the point more generally, the aim of the analogy with physics was to show how we could remove the apparent paradoxes and absurdities by showing that they were just as apparent but as unreal as in the physics case. We see that there is no absurdity in supposing that we can be going both 5 and 60 miles an hour at the same time, once we see that our speed is relative to different coordinate systems. But the analogy between physics and meaning fails. Even after we have relativized meaning, we are still left with the same absurdities we had before.

Why does the analogy break down? In physics the position and motion of a body consist entirely in its relations to some coordinate system; but there is more to meaning than just the relations that a word has to the language of which it is a part; otherwise the question of translation could never arise in the first place. We can't detach the specific motion or position of an object from a reference to a specific coordinate system and translate it into another system in the way we can detach a specific meaning from a specific linguistic system and find an expression that has that very meaning in another linguistic system. Of course, a word means what it does only relative[12] to a language of which it is a part, but the very relativity of the *possession* of meaning presupposes the nonrelativity of the *meaning* possessed. This has no analogue in the relativity of physical position and motion.

Someone might object that I seem to be assuming the very "myth of the museum" that Quine is challenging, the view that there exists a class of mental entities called "meanings." But my point is neutral between the various theories of meaning. Let meaning be a matter of ideas in the head à la Hume, dispositions to behavior à la Quine, uses of words à la Wittgenstein, or intentional capacities à la me. It doesn't matter for this point. Whatever meaning is, we need to distinguish the true thesis that a word has the particular meaning it has only relative to a language from the false thesis that the meaning itself is relative to a language. Indeed, we are now in a position to state the argument in a way that is independent of any particular

[12] I argue elsewhere that the functioning of a speaker's meaning is also relative to a whole Network of intentional states and a Background of preintentional capacities. I believe that this relativity is vastly more radical than has been generally appreciated and, indeed, more radical than Quine's indeterminacy thesis, but it is irrelevant to this part of the indeterminacy dispute. [See my *Intentionality: An Essay in the Philosophy of Mind* (New York: Cambridge, 1983), chaps. 1 and 5.]

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

theory of meaning: grant me that there is a distinction between meaningful and meaningless phonetic sequences (words). Thus, in English, 'rabbit' is meaningful, 'flurg' is meaningless. Such remarks are always made relative to a language. Perhaps in some other language 'flurg' is meaningful and 'rabbit' is meaningless. But if 'rabbit' is meaningful in English and 'flurg' is meaningless, there must be some feature that 'rabbit' has in English which 'flurg' lacks. Let's call that feature its *meaning*, and the class of such features of words we can call *meanings*. Now, from the fact that 'rabbit' has the particular feature it has relative to English, it does not follow that the feature, its meaning, can exist only relative to English. Indeed, the question whether 'rabbit' has a translation into another language is precisely the question whether in the other language there is an expression with that very feature. The analogy between relativity in physics and semantics breaks down because there are no features of position and motion except relations to coordinate systems. And Quine's argument is a reductio ad absurdum because it shows that the totality of dispositions to speech behavior is unable to account for distinctions concerning the feature, meaning, which we know independently to exist, the distinction between the meaning of 'rabbit' and that of 'rabbit stage', for example. You cannot avoid the reductio by calling attention to the fact that 'rabbit' has the feature, its meaning, only relative to English, because the reductio is about the feature itself, and the feature itself is not relative to English.

My aim so far has not been to refute extreme linguistic behaviorism, but to show:

> First, the thesis of the indeterminacy of translation is just as well (indeed, I think better) construed as a reductio ad absurdum of the premises from which it was derived as it is construed as a surprising result from established premises.
>
> Second, the theory of the relativity of ontology does not succeed in answering the apparent absurdities that the thesis of indeterminacy and inscrutability leads us into.

What about refuting linguistic behaviorism on its own terms? There have been so many refutations of behaviorism in its various forms that it seems otiose to repeat any of them here. But it is worth pointing out that Quine's argument has the form of standard and traditional refutations of behaviorism. We know from our own case, from the first-person case, that behaviorism is wrong, because we know that our own mental phenomena are not equivalent to dispositions to behavior. Having the pain is one thing, being disposed to exhibit pain behavior is another. Pain behavior is insufficient to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

account for pain, because one might exhibit the behavior and not have the pain, and one might have the pain and not exhibit it. Analogously, on Quine's argument, dispositions to verbal behavior are not sufficient to account for meanings, because one might exhibit behavior appropriate for a certain meaning, but that still might not be what one meant.

If someone has a new theory of the foundations of mathematics and from his new axioms he can derive that 2 + 2 = 5, what are we to say? Do we say that he has made an important new discovery? Or do we say, rather, that he has disproved his axioms by a reductio ad absurdum? I find it hard to imagine a more powerful reductio ad absurdum argument against behaviorism than Quine's indeterminacy argument, because it denies the existence of distinctions that we know from our own case are valid.

IV

I have tried to show how the doctrines of indeterminacy and inscrutability depend on the special assumptions of behaviorism and that, consequently, the results can equally be taken as a refutation of that view. But now an interesting question arises. Why do philosophers who have no commitment to behaviorism accept these views? I will consider Donald Davidson, because he accepts the doctrine of indeterminacy while explicitly denying behaviorism. Davidson takes the frankly intentionalistic notion of "holding a sentence true" (i.e., believing that it is true) as the basis on which to build a theory of meaning. What then is the area of agreement between him and Quine which generates the indeterminacy? And what does he have to say about the "quandary" that Quine faces? How does he deal with the first-person case? Davidson answers the first question this way:

> The crucial point on which I am with Quine might be put: all the evidence for or against a theory of truth (interpretation, translation) comes in the form of facts about what events or situations in the world cause, or would cause, speakers to assent to, or dissent from, each sentence in the speakers' repertoire (*op. cit.*, 230).

That is, as long as the unit of analysis is a whole sentence and as long as what causes the speaker's response is an objective state of affairs in the world—whether the response is assent and dissent, as in Quine, or holding a sentence true, as in Davidson—Davidson agrees with Quine about the indeterminacy thesis. (There are some differences about the extent of its application.)

But how exactly does the argument work for Davidson? How does Davidson, who rejects behaviorism, get the result that reference is inscrutable? I believe a close look at the texts suggests that he does

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

accept a modified version of Quine's conception of an empirical theory of language. Though he accepts an intentionalistic psychology, he insists that semantic facts about the meanings of utterances must be equally accessible to all the participants in the speech situation, and thus for him the first-person case has no special status.

Quine grants us an apparatus of stimuli and dispositions to verbal response. Davidson grants us conditions in the world (corresponding to Quine's stimuli), utterances, and the psychological attitude of "holding true," directed at sentences. But, since the unit of empirical test is still the sentence, as opposed to parts of the sentence, and since different schemes of interpreting sentences in terms of parts of sentences can be made consistent with the same facts about which sentences a speaker holds true and under what conditions the speaker holds those sentences true, Davidson claims we still get inscrutability. The basic idea is that there will be different ways of matching up objects with words, any number of which could equally well figure in a truth theory that explained why a speaker held a sentence true.

The puzzle about Davidson is that, if you set out the argument as a series of steps, it doesn't follow that there is inscrutability *unless* you add an extra premise concerning the nature of an empirical theory of language. Here are the steps:

(1) The unit of empirical analysis in radical interpretation is the sentence (as opposed to subsentential elements).
(2) The only empirical evidence for radical interpretation is the fact that speakers "hold true" certain sentences in certain situations.
(3) There are alternative ways of matching words with objects which are inconsistent, but any number of which could equally well explain why a speaker held a sentence true.

But these three do not entail any inscrutability or indeterminacy about what the speaker actually meant and what he is referring to. For that you need an extra premise. What is it? I believe that it amounts to the following:

(4) All semantic facts must be publicly available to both speaker and hearer. If the interpreter cannot make a distinction on the basis of *public,* empirical evidence, then there is no distinction to be made.

Here is one of his examples: if everything has a shadow, then in a circumstance in which a speaker holds true the sentence 'Wilt is tall', we can take 'Wilt' to refer to Wilt and 'is tall' to refer to tall things, or we can with equal empirical justification take 'Wilt' to refer to the shadow of Wilt and 'is tall' to refer to the shadows of tall things. The first theory tells us that 'Wilt is tall' is true iff Wilt is tall. The second

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

theory tells us that 'Wilt is tall' is true iff the shadow of Wilt is the shadow of a tall thing.

Davidson summarizes the argument thus:

> The argument for the inscrutability of reference has two steps. In the first step we recognize the empirical equivalence of alternative reference schemes. In the second step we show that, although an interpreter of the schemer can distinguish between the schemer's schemes, the existence of alternative schemes for interpreting the schemer prevents the interpreter from uniquely identifying the reference of the schemer's predicates, in particular his predicate 'refers' (whether or not indexed or relativized). *What an interpreter cannot on empirical grounds decide about the reference of a schemer's words cannot be an empirical feature of those words.* So those words do not, even when chosen from among arbitrary alternatives, uniquely determine a reference scheme (235; my italics).

In order to understand this argument it is crucial to see that it rests on the special assumption I mentioned about the nature of an empirical account of language and about the public character of semantics. From the mere fact that alternative reference schemes are consistent with all the *public* empirical data it simply doesn't follow by itself that there is any indeterminacy or inscrutability. Indeed, this is simply the familiar undetermination thesis all over again: different hypotheses will account equally for the speaker's "hold true" attitudes, but, all the same, one of the hypotheses may be right about exactly what he meant by his words while another hypothesis may be wrong. In order to get the result of inscrutability, an additional premise is needed: since language is a public matter, all the facts about meaning must be public facts. Meaning is an "empirical" matter, and what is empirical about language must be equally accessible to all interpreters. Only given this assumption, this special conception of what constitutes the "empirical" and "public" character of language, can the argument be made to go through.

In order to deepen our understanding of what is going on here, let us contrast the common-sense account of the speech situation with Davidson's account. On the common-sense account, when I make the assertion, "Wilt is tall," by 'Wilt' I refer to Wilt, and by 'is tall' I mean: is tall. When I say "Wilt," I make no reference explicitly or implicitly to shadows, and, similarly, when I say "is tall," I make no reference to shadows. Now these are just plain facts about me. They are not theoretical hypotheses designed to account for my behavior or my "hold-true" attitudes. On the contrary, any such theory has to start with facts such as these. But, on Davidson's view, there is no empirical basis for attributing these different intentional states to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

me. Since all the empirical facts we are allowed to use are facts about what sentences I hold true and under what (publicly observable) conditions, there is no way to make the distinctions that our common-sense intuitions insist on. As with behaviorism, different and inconsistent interpretations at the subsentence level, at the level of words and phrases, will all be consistent with all the facts about what sentences I hold true under what conditions. But now it begins to look as if Davidson's version of inscrutability might also be a reductio ad absurdum of his premises, just as Quine's account was a reductio ad absurdum of behaviorism.

Before we draw any such conclusion, let us first see how Davidson deals with the obvious objection that is suggested by the common-sense account: since we do know in our *own* use of language that we are referring to Wilt, for example, and not to Wilt's shadow, and since what we seek in understanding another person is precisely what we already have in our own case, namely (more or less) determinate senses with determinate references, why should anyone else's references and senses be any less determinate than our own? Of course, in any given case I might get it wrong. I might suppose someone was referring to Wilt when really it was the shadow he was talking about. But that is the usual underdetermination of hypotheses about other minds from publicly available evidence. It does not show any form of inscrutability. What, in short, does Davidson say about the "quandary" that Quine faces, the first-person case?

> Perhaps someone (not Quine) will be tempted to say, 'But at least the speaker knows what he is referring to.' One should stand firm against this thought. The semantic features of language are public features. What no one can in the nature of the case figure out from the totality of the relevant evidence cannot be a part of meaning. And since every speaker must, in some dim sense at least, know this, *he cannot even intend to use his words with a unique reference for he knows that there is no way for his words to convey the reference to another* (235; my italics).

Quine tries to avoid the quandary by an appeal to relativity, but on Davidson's view there really isn't any quandary in the first place. Semantic features are public features, and since the public features are subject to the indeterminacy, *there is no such thing as unique reference.* Furthermore, "in some dim sense" I must know this; so *I can't even intend to refer to rabbits as opposed to rabbit parts, and I can't intend to refer to Wilt as opposed to Wilt's shadow.*[13]

[13] Kirk Ludwig has pointed out to me that this seems to lead to a pragmatic paradox, since it looks as if, in order to state the thesis, we have to specify distinctions that, the thesis says, cannot be specified.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Now, I believe this is a very strange view to hold, and I propose to examine it a bit further. First of all, let us grant that, for "public" languages such as French and English, there is at least one clear sense in which semantic features are, indeed, public features. I take it all that means is that different people can understand the same expressions in the same way in French and English. Furthermore, let us grant, at least for the sake of argument, that the public features are subject to underdetermination in at least this sense: I could give different but inconsistent interpretations of someone's words, all of which would be consistent with all of the actual and possible evidence I had about which sentences he held true. Now what follows? In our discussion of Quine's view we saw that indeterminacy, as opposed to underdetermination, is a consequence only if we deny mentalism from the start; it is not a consequence of underdetermination by itself. But, similarly, on Davidson's view the indeterminacy follows only if we assume from the start that different semantic facts must necessarily produce different "publicly observable" consequences. Only given this assumption can we derive the conclusion that speaker's meaning and reference are indeterminate and inscrutable. But, I submit, we know quite independently that this conclusion is false, and, therefore, the premises from which it is derived cannot all be true. How do we know the conclusion is false? We know it because in our own case we know that we mean, e.g., Wilt as opposed to Wilt's shadow, rabbit as opposed to rabbit stage. When I seek to understand another speaker, I seek to acquire in his case what I already have for my own case. Now, in my own case, when I understand myself, I know a great deal more than just under what external conditions I hold what sentences true. To put it crudely: in addition, I know what I mean. Furthermore, if another person understands me fully, he will know what I mean, and this goes far beyond just knowing under what conditions I hold what sentences true. So, if his understanding me requires much more than just knowing what sentences I hold true under what conditions, then my understanding him requires much more than knowing what sentences he holds true under what conditions. Just knowing his "hold true" attitudes will never be enough for me fully to understand him. Why should it be? It would not be enough for me to understand me; and since, to repeat, what I need to acquire in his case is what I already have in my own case, I will need more than just these attitudes.

But what about Davidson's claim that what an interpreter cannot figure out from the totality of the relevant evidence cannot be part of meaning? Well, it all depends on what we are allowed to count as "figuring out from the totality of the relevant evidence." On the common-sense account, I do figure out from the relevant "evi-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

dence" that by 'Wilt' you mean Wilt and not Wilt's shadow, and the "evidence" is quite conclusive. How does it work? In real life I understand the speech of another not only within a Network of shared assumptions, but more importantly against a Background of nonrepresentational mental capacities—ways of being and behaving in the world which are both culturally and biologically shaped and which are so basic to our whole mode of existence that it is hard even to become aware of them (see my *Intentionality, op. cit.*, ch. 5). Now, given the Background, it will, in general, be quite out of the question that, when you say in English, "Wilt is tall" or "There goes a rabbit," you could with equal justification be taken to be talking about Wilt's shadow or rabbit stages. We get that surprising result only if we forget about real life and imagine that we are trying to understand the speech of another by constructing a "theory," using as "evidence" only his "hold true" attitudes directed toward sentences or his dispositions to make noises under stimulus conditions. Language is indeed a public matter, and, in general, we can tell what a person means if we know what he says and under what conditions he says it. But this certainty derives not from the supposition that the claim about what he means must be just a summary of the (publicly available) evidence; it is rather the same sort of certainty we have about what a man's intentions are from watching what he is doing. In both cases we know what is going on because we know how to interpret the "evidence." And in both cases the claims we make go beyond being mere summaries of the evidence, in a way that any claim about "other minds" goes beyond being a summary of the "public" evidence. But the fact that the interpretation of the speech of another is subject to the same sort of underdetermination[14] as any other claim about other minds does not show either that there is any indeterminacy or that we cannot, in general, figure out exactly what other people mean from what they say.

I conclude that our reaction to Davidson's version should be the same as our reaction to Quine's: in each case the conclusion of the argument is best construed as a reductio ad absurdum of the premises. Davidson's view is in a way more extreme than Quine's because

[14] Here is an example of such undetermination from real life. Until he was in middle age, a friend of mine thought that the Greek expression *hoi polloi* as used in English meant the elite of rich people, but that it was characteristically used ironically. Thus, if he saw a friend in a low-class bar he might say, "I see you have been hobnobbing with the hoi polloi." Since he spoke ironically and interpreted other people as speaking ironically, there were no behavioral differences between his use and the standard use. Indeed, he might have gone his whole life with this semantic eccentricity undetected. All the same, there are very definite facts about what he meant.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

he holds a view which is, I believe, literally incredible. Plugging in the first-person example to what he literally says, Davidson holds that what no external observer can decide from external evidence cannot be part of what I mean. Since such observers can't decide between inconsistent interpretations, and since I must, in some dim sense at least, know this, I cannot even intend to use 'rabbit' to mean rabbit as opposed to rabbit stage or undetached rabbit part, for I know there is no way for my words to convey this reference to another. This does not seem to me even remotely plausible. I know exactly which I mean, and though someone might get it wrong about me, just as I might get it wrong about him, the difficulty is the usual "other-minds problem" applied to semantics.

V

In any discussion like this there are bound to be issues much deeper than those which surface in the actual arguments of the philosophers involved. I believe that the deepest issue between me on the one hand and Davidson and Quine on the other concerns the nature of an empirical theory of language.

Both Quine and Davidson adopt the thought experiment of "radical translation" as a model for building an account of meaning. In radical translation an interpreter or translator tries to understand speakers of a language of which he has no prior knowledge whatever. On Davidson's view, "all understanding of the speech of another involves radical interpretation."[15] But the model of an unknown foreign language enables us to make more precise what sorts of assumptions and evidence we need to interpret someone else's speech.

Notice that the model of radical translation already invites us, indeed forces us, to adopt a third-person point of view. The question now becomes, How would *we* know the meaning of the utterances of some *other* person? And the immediate difficulty with that way of posing the question is that it invites confusion between the epistemic and the semantic; it invites confusion between the question, *How* do you know? and the question, *What* is it that you know *when* you know? But the linguistically relevant facts must be the same in the questions, What is it for me to understand another person when he says "It's raining"? and What is it for me to understand myself when I say "It's raining"? since, to repeat, what I have when I understand him is exactly what he has when he understands me. But then I already understand me; so anything I can learn from studying his case I could learn from studying my case.

[15] "Radical Interpretation," *Dialectica*, XXVII (1973): 313–328, reprinted in *Inquiries into Truth and Interpretation*, pp. 125–139.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Still, the thought experiment of radical translation can be very useful in semantic theory because it focuses the question of how we communicate meaning from one speaker to another. The difficulty is that both Quine and Davidson set further constraints on the task of radical translation than those which any field linguist would in fact employ. I have twice watched the linguist Kenneth L. Pike[16] perform the "monolingual demonstration" where he begins to construct a translation of a totally alien language into English. And it seems quite clear to any observer of Pike that he does not confine his conception of translation to that described by Davidson and Quine. For example, Pike does not confine his investigation to matching verbal behavior and sensory stimuli in the manner of Quine, nor does he confine it to hold-true attitudes in the manner of Davidson. Rather, he tries to figure out what is going on in the mind of the native speaker, even at the level of particular words. And he can do this because he presupposes that he shares with the speaker of the exotic language a substantial amount of Network and Background (see fn 12 above).

Now granted that the thought experiment of radical interpretation is useful in understanding the notion of communication, why shouldn't the problem of radical interpretation be posed in common-sense mentalistic terms? Why should we place on it the further behavioristic or "empirical" constraints that Quine and Davidson so obviously do? Quine's writings contain scattered remarks of the following sort: "Our talk of external things, our very notion of things, is just a conceptual apparatus that helps us to foresee and control the triggering of our sensory receptors in the light of previous triggering of our sensory receptors. The triggering, first and last, is all we have to go on."[17]

Such a remark has the air of discovery, but I believe it simply expresses a preference for adopting a certain level of description. Suppose one substituted for the phrase "triggering of our sensory receptors" in this paragraph, the phrase "the movement of molecules." One could then argue that the movement of molecules, first and last, is all we have to go on. Both the "movement of molecules" version and the "sensory receptors" version are equally true and equally arbitrary. In a different philosophical tradition, one might also say that all we have to go on, first and last, is the thrownness (*Geworfenheit*) and the foundedness (*Befindlichkeit*) of Dasein in the lifeworld (*Lebenswelt*). Such remarks are characteristic of philo-

[16] Pike's work appears to be the original inspiration for the idea of radical translation (see Quine, *Word and Object*, p. 28).
[17] *Theories and Things*, p. 1.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sophy, but it is important to see that what looks like a discovery can equally be interpreted as simply the expression of preference for a certain level of description over others. The three choices I gave are all equally interpretable as equally true. How do we choose among them? I believe that all three—sensory receptors, molecules, and Dasein—are insufficient levels of description for getting at certain fundamental questions of semantics. Why? Because the level of semantics that we need to analyze also involves a level of intentionality. Semantics includes the level at which we express beliefs and desires in our intentional utterances, at which we mean things by sentences and mean quite specific things by certain words inside of sentences. Indeed, I believe that the intentionalistic level is already implicit in the quotation from Quine when he uses the expressions 'foresee' and 'control'. These convey intentionalistic notions, and, on Quine's own version of referential opacity, they create referentially opaque contexts. No one, with the possible exception of a few neurophysiologists working in laboratories, tries to foresee and control anything at the level of sensory receptors. Even if we wanted to, we simply don't know enough about this level. Why then in Quine do we get this round declaration that all we have to go on is the stimulation of the sensory receptors? I think it rests on a resolute rejection of mentalism in linguistic analysis, with a consequent insistence on having a third-person point of view. Once you grant that a fundamental unit of analysis is intentionality, then it seems you are forced to accept the first-person point of view as in some sense epistemically different from the point of view of the third-person observer. It is part of the persistent objectivizing tendency of philosophy and science since the seventeenth century that we regard the third-person objective point of view as preferable to, as somehow more "empirical" than, the first-person, "subjective" point of view. What looks then like a simple declaration of scientific fact—that language is a matter of stimulations of nerve endings—turns out on examination to be the expression of a metaphysical preference and, I believe, a preference that is unwarranted by the facts. The crucial fact in question is that performing speech acts—and meaning things by utterances—goes on at a level of intrinsic first-person intentionality. Quine's behaviorism is motivated by a deep antimentalistic metaphysics which makes the behaviorist analysis seem the only analysis that is scientifically respectable.

A similar though more subtle form of rejection of the first-person point of view emerges in Davidson's writings in a number of places. Davidson tacitly supposes that what is empirical must be equally and publicly accessible to any competent observer. But why should it be?

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

It is, for example, a plain empirical fact that I now have a pain, but that fact is not equally accessible to any observer. In Davidson, the crucial claims in the passages I quoted are where he says, "What an interpreter cannot on empirical grounds decide about the reference of a schemer's words cannot be an empirical feature of those words"; and prior to that where he claims, "What no one can in the nature of the case figure out from the totality of the relevant evidence cannot be a part of meaning." Both of these have an air of truism, but in actual usage they express a metaphysical preference for the third-person point of view, a preference which is assumed and not argued for; because, as in Quine's case, it seems part of the very notion of an empirical theory of language, an obvious consequence of the fact that language is a public phenomenon. What Davidson says looks like a tautology: What can't be decided empirically isn't empirical. But the way he uses this is not as a tautology. What he means is: What can't be conclusively settled on third-person objective tests cannot be an actual feature of language as far as semantics is concerned. On one use "empirical" means: subject to objective third-person tests. On the other use it means: actual or factual. There are then two different senses of "empirical"; and the argument against the first-person case succeeds only if we assume, falsely, that what isn't conclusively testable by third-person means isn't actual. On the other hand, once we grant that there is a distinction between the public evidence available about what a person means and the claim that he means such and such—that is, once we grant that the familiar underdetermination of evidence about other minds applies to semantic interpretation—there is no argument left for inscrutability.

The rival view that is implicit in my argument is this. Language is indeed public; and it is not a matter of meanings-as-introspectable-entities, private objects, privileged access, or any of the Cartesian paraphernalia. The point, however, is that, when we understand someone else or ourselves, what we require—among other things—is a knowledge of intentional contents. Knowledge of those contents is not equivalent to knowledge of the matching of public behavior with stimuli nor to the matching of utterances with conditions in the world. We see this most obviously in the first-person case, and our neglect of the first-person case leads us to have a false model of the understanding of language. We think, mistakenly, that understanding a speaker is a matter of constructing a "theory," that the theory is based on "evidence," and that the evidence must be "empirical."

JOHN R. SEARLE
University of California/Berkeley

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE EPISTEMOLOGY OF FIRST-PERSON REFERENCE*

IT is a general requirement on one's over-all approach to semantics that it be harmonious with one's epistemological views. But, even when the two have been coordinated at a general level, the first-person singular requires special attention. For, although the pronoun 'I' in English is governed by a simple rule—'I' refers to the speaker—the rule can be reformulated in such a way as to reflect what looks like a special epistemic relation. 'I' seems to express a sort of self-consciousness or reflexive self-awareness in virtue of which a speaker knows that in her uses of the term 'I' she refers to herself, not under any particular description such as 'the speaker', but simply as *herself*. We can reformulate the rule to bring this out as follows: the first-person singular pronoun is that word in a language which a speaker uses to refer to herself, where the term 'herself' is an emphatic reflexive. The emphatic reflexive is just a way of rendering in the third person what is essentially first-personal and, hence, irreducible to any other terms.[1]

Elizabeth Anscombe insists, in light of this feature of 'I' (along with others which I shall elaborate below), that the simple rule that 'I' refers to the speaker is not by itself adequate for the purposes of a Fregean account of 'I':

> We seem to need a sense to be specified for this quasi-name "I." To repeat the Frege point: we haven't got this sense just by being told which object a man will be speaking of [e.g., the speaker], whether he knows it or not, when he says "I." Of course that phrase "whether he knows it or not" seems highly absurd. His use of "I" surely guarantees that he does

* I wrote the first draft of this paper while participating in a Summer Seminar for College Teachers sponsored by the National Endowment for the Humanities. I am grateful to the Endowment, as well as to Sydney Shoemaker, who ran the seminar and who provided many useful comments both during and after the seminar. I would also like to thank Akeel Bilgrami, Jennifer Church, Donald Davidson, Leon Galis, Isaac Levi, John Searle, and Stephen White for discussing an earlier version with me.

[1] The term 'emphatic reflexive' I owe to Roderick Chisholm; see his *The First Person* (Minneapolis: Minnesota UP, 1981). The impossibility of rendering or reporting first-person singular modes of thought and speech in the third person reductively (i.e., without the emphatic reflexive) has been the starting point for a growing literature about 'I' which, apart from Chisholm's book, includes: Hector-Neri Castañeda, "He: A Study in the Logic of Self-Consciousness," *Ratio*, VIII, 2 (December 1966): 130–157; G. E. M. Anscombe, "The First Person," in Samuel Guttenplan, ed., *Mind and Language* (New York: Oxford, 1975), pp. 45–65; John Perry, "The Problem of the Essential Indexical," *Noûs*, XIII, 1 (March 1979): 3–21; David Lewis, "Attitudes *De Dicto* and *De Se*," *Philosophical Review*, LXXXVIII, 4 (October 1979): 513–543.

0022-362X/87/8403/0147$02.10 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

> know it! But we have a right to ask *what* he knows: if "I" expresses a way its object is reached by him, what Frege called an "Art des Gegebenseins," we want to know what that way is and how it comes about that the only object reached in that way by anyone is identical with himself (*op. cit.*, 48).

We can view Anscombe's Fregean requirement as a demand for an epistemological framework that provides the backdrop against which we can set the simple rule governing 'I' in order for us to gain a complete understanding of the distinctive features of the first person (its irreducibility being only one of them). Even those who do not adhere as Anscombe does to a strict Fregean conception of reference may want to meet the challenge she couches in Fregean terms. Since it is not at all obvious that every referring expression expresses one and only one way of "reaching" its object, we might be perfectly content to settle for something less than a strict Fregean sense for 'I'.[2] But, whether strict or relaxed, there are widely known reasons for supposing that all reference must involve some way of identifying or individuating the referent. Only then can we give substance to the idea that we refer to one thing and not another.

The upshot of Anscombe's discussion then is this: the distinctive features of the first person invite the idea that 'I' refers directly, i.e., in a manner unmediated by any criteria for identity or individuating concepts and beliefs. If as Fregeans we were to accept this idea, our account of first-person-singular reference should commit us to a corresponding epistemological thesis positing some kind of direct self-awareness in terms of which we could explain the possibility of direct self-reference as a case apart from the ordinary. The situation is different for subscribers to the causal theory of reference. Since they already view proper names and natural-kind terms as directly referring, they could perhaps assimilate first-person reference to

[2] My commitment to Frege's distinction between sense and reference is not, as I've called it, "strict," but rather at the level of very general principle. Attempts to elucidate the notion of sense as a sort of definite description fall short, since most proper names are not associated with any one description that serves to (or is adequate to) individuate the referent. The cluster theory [see John Searle, "Proper Names," *Mind*, XLVII, 266 (April 1958): 166–173] makes a well-known modification of the strict Fregean notion by allowing for a list of descriptions for a proper name such that one (or more) of the descriptions must be appealed to in (or again, is adequate for) the identification of the referent. I myself prefer that Searle's modification be situated in a theory of belief for a speaker. Holistic considerations support the idea that the point of Frege's distinction can be captured within a coherentist account of reference which insists on a background of interlocking concepts and beliefs that together (albeit in a loose and contextual way) illuminate the difference between the referent of a term and other objects. In my account I demand no more than this weak conception.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

their view and thereby salvage the idea of first-person reference as direct, without appeal to a kind of direct access to the self which is *sui generis*. However, anyone who has Fregean sentiments (thereby repudiating on principle the idea of direct or unmediated reference) faces a choice. Either first-person reference must be accepted as direct and hence as an exception to the over-all Fregean stance requiring special epistemological backing, or first-person reference must be assimilated to the Fregean model. Those who choose the latter course will want to meet Anscombe's challenge by explaining the distinctive features of the first person without appeal to any notion of direct epistemic access; for this notion *cannot* in any way serve to explain *indirect* or mediated first-person reference.

Over the next many sections I shall argue that Anscombe's challenge can be met. She disagrees, contending that no sense can be found for 'I' and concluding that 'I' is not really a referring term. I shall thus be arguing against Anscombe that 'I' does refer, but with her (against the direct-reference view) that 'I' must meet Frege's general desideratum for all referring expressions. To the extent that I succeed in showing that it does meet it, I will claim to have sketched the rudiments of a satisfactory epistemology of first-person reference.

As she elaborates her case that 'I' does not refer, Anscombe fastens on two special features of the first person, to which I will add a third which would seem to provide additional support for her case. Taken together, these features encourage, and have been taken to encourage, the direct-reference view. First, uses of 'I' are accompanied by what Anscombe calls "referential guarantees," which are of two sorts. In all its uses the referent of 'I' is guaranteed to exist. But, more importantly, it is extremely rare (though not impossible) that a speaker should misidentify something that is not the referent (that is, not herself) as the referent of her use(s) of 'I'. Second, there is the already noted irreducibility of 'I' to any terms that do not contain the emphatic reflexive. In particular, 'I' is irreducible to any name or description; whereas any name or description that refers to a subject may be such that the subject fails to know that it refers to him, he cannot fail to know that he refers to himself with 'I'. Connected with this claim is a third feature of the first person, viz., that first-person thought plays a special role in bringing about intentional action which no other (non-first-personal) mode of thought about oneself can play.[3] These features of the first person add up to an invitation to

[3] Perry, *op. cit.*, recounts a story of following a trail of sugar on a supermarket floor. He believed that the shopper with the torn sack should be stopped, and his

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

posit direct first-person reference, supported by an epistemological corollary about direct self-awareness because: (a) the referential guarantees surrounding 'I' (or rather just the second guarantee: 'I' 's immunity to error through misidentification) indicate that our referential grip on ourselves is somehow tighter than it is on other objects; (b) the irreducibility of 'I' suggests that our grip is tighter because it isn't mediated by sense; and (c) the special psychological role of the first person leads us to entertain the possibility that rational agency involves just the sort of self-consciousness that could account for the capacity of speakers to refer to themselves unmediatedly. An example of this capacity is the one Roderick Chisholm (*op. cit.*) and David Lewis (*op. cit.*) appeal to in their accounts of first-person attitudes: the capacity to self-ascribe properties directly, and an alternative is provided by Stephen Schiffer,[4] who proposes that we invoke Bertrand Russell's notion of acquaintance.

My task, then, is to account for all these distinctive features of the first person described above without recourse to any notion of a special direct self-awareness. I am not the first to suggest that we can do this. For instance, Steven Boer and William Lycan[5] argue that 'I' is not semantically unique, but rather is pragmatically distinctive. They are, however, roughly in the camp of the causal theorists, their main thesis being that first-person attitudes are reducible to *de re* attitudes of oneself. I take it that the semantic corollary is that we should see 'I' (in each of its uses, for it is an indexical expression) as rigidly designating the speaker. There is also John Perry's claim (*op. cit.*) that finding a sense for 'I' is beside the point and that what is truly distinctive about first-person attitudes has nothing to do with reference to oneself. Rather, we should concentrate exclusively on the special role of these attitudes in motivating (and hence explaining) actions. But neither of these suggestions fully meets Anscombe's

---

belief was, unbeknownst to him, a belief about himself. Only after he realized that *he himself* was the shopper could he act appropriately on his belief. I should perhaps mention that other demonstratives also play this special motivational role: Trying to find Magdalen College, I was told to go left at Carfax. I expected some sort of marker, but there was none, and so on arriving at Carfax I failed to turn left, not knowing of course that *that* was Carfax. However, pronouns like 'this' and 'that' are not accompanied by the referential guarantees that preoccupy Anscombe in her treatment of 'I'. 'Here' and 'now' do exhibit them and, therefore, equally stand in need of an account of the sort I am trying to provide for 'I' in this paper (see fn 18 below for more discussion of this point). The difference between the two is not, then, over the fact that all three features are present and need accounting for, but rather that 'I', unlike 'here' and 'now', is caught up in the question of self-consciousness and self-identity.

[4] "The Basis of Reference," *Erkenntnis*, XIII, 2 (August 1978): 171–206.

[5] "Who Me?," *Philosophical Review*, LXXXIX, 3 (July 1980): 427–466.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

challenge; for neither brings out how reference to oneself depends on in some way *individuating* oneself in precisely the way that reference in general depends on individuation (i.e., on something that does the job of Fregean sense). Unlike those authors, therefore, I will try to account for these distinctive features in terms of the very features that meet the Fregean desideratum.

My discussion is organized in terms of the distinctive features of the first person outlined above. In the first section I examine the usual sources of referential failure and suggest how we can account for the immunity of 'I' from these failures without giving up the Fregean commitments. I propose that it is in fact constitutive of our referring in the first person that we be especially clear about our identities. In section II I consider a possible objection to my proposal which makes a connection between the referential guarantees associated with the first person and the conceivability of a purely inner use of 'I' not associated even loosely with criteria for the identity of the subject. Then, having addressed that objection, I turn in section III to the irreducibility of 'I' to any other referring expression. I argue that the irreducibility of 'I' to any name or description does not imply that 'I' refers independently of individuating descriptive beliefs and, hence, does not provide any reason for thinking that first-person reference is direct in some ineffable, distinctively first-personal way. Section IV is devoted to showing that not only do we not need to posit any sort of direct reference or self-awareness in order to understand the special motivational force of first-person attitudes, but, on the contrary, we must posit precisely the sort of background of descriptive beliefs which gives substance to the idea that a subject knows which thing she is when she uses 'I'. Finally, in section V I show that there is an intimate connection between this last point and my proposal in section I. I also place my account of first-person reference in a broader philosophical context in which the self-consciousness that we express with the first person is linked with our status as social, communicative beings.

I

Referential failures occur when a speaker's referential intentions are not in synch with the rules that govern the use of a referring expression. The result is that the speaker's utterance is either inappropriate or straightforwardly false. These failures are of two sorts, which I will call, respectively, *failure of existence* and *error through misidentification.*[6] In the first sort a speaker falsely believes that there is some-

[6] The latter phrase is Sydney Shoemaker's; see his "Self-reference and Self-awareness," this JOURNAL, LXV, 19 (Oct. 3, 1968): 555–567, p. 556.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

thing that meets the requirements of the referring expression, when nothing does. This is often a contextual matter, as, for instance, when a speaker says, pointing to an empty cookie jar, "Those must be stale by now." The speaker clearly believes that there are some cookies in the jar and intends to refer to them. It is a condition on successful reference in this context that the belief be true. But it is false, and so the speaker has literally failed to refer. Referential error through misidentification is also the product of false beliefs on the part of a speaker. In this case it is not a false existential belief (of the form 'there is a thing that is *F*'), but a false belief about the identity of a given object (having roughly the form 'the *F* is the *G*' or '*a* is *b*'). The belief leads the speaker to use a referring expression in order to refer to a given object even though its semantically determined referent is not that object. This is a familiar phenomenon, and one which led Keith Donnellan to distinguish between referential and attributive uses of referring expressions.[7] Here is an example: Henry Brown is introduced to a woman named "Rebecca Thompson" at a college function and falsely believes that he is being introduced to the dean. He notices that she leaves the function at the earliest possible moment and says, "I see that the dean doesn't feel obliged to stay for the president's speech." He has said something that is straightforwardly false, because the dean is in fact right at the college president's side. The only person who has left is Rebecca Thompson. An interpreter who knew this would naturally account for Henry's false assertion by crediting him with the false belief that Rebecca Thompson is the dean. We might not want to say in such a situation (as we did in the case of failed existence) that the speaker has literally failed to refer. In fact Donnellan introduced the notion of an attributive use of a referring expression in order to make room for the idea that reference to an object can still go through despite the inappropriateness of the expression a speaker uses in referring to the object. In this way we can interpret Henry as referring to Rebecca Thompson despite his ill-chosen word and as saying something true about her (rather than something false about the dean).[8] I will, however, continue to call this a kind of referential failure, mainly in order to maintain a symmetry with Anscombe's phrase "referential guarantee" which precludes both kinds of error.

There is no gainsaying that 'I' is immune from the first kind of

[7] "Reference and Definite Descriptions," *Philosophical Review*, LXXV, 3 (July 1966): 281–304.

[8] Shoemaker suggests that we should see such cases as cases of divided reference, and then we might say that Henry has accomplished two things, i.e., saying something true of Rebecca Thompson and false of the dean.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

referential failure. 'I' refers to the speaker, and any speaker who has mastered the term knows this rule. Presumably, any speaker also knows that she is speaking when she uses 'I', and if she knows this she surely knows that there is a speaker. Which is just to say that, in any use of 'I', the referent (the speaker) is guaranteed to exist and, moreover, that speakers know this. This guarantee of existence does not, however, cast any light on *how* first-person reference is achieved or whether it is direct. If immunity from referential failure is supposed to lend plausibility to the claim that first-person reference is direct, it must then be immunity from the second kind.

At first glance it is hard to see how referential error through misidentification could plague uses of 'I' any more than failure of existence can. For if a speaker knows that 'I' refers to the speaker, and if a speaker knows that she is speaking whenever she uses 'I', how can she intend to refer to anyone but herself when she uses 'I'? That would seem tantamount to saying that she can utter 'I' and believe that someone who is not herself has uttered it. We can actually make some limited sense of this last possibility, though my way of putting things obscures it. Let me describe, therefore, a situation in which we might be prepared to grant that a use of the first person is accompanied by error through misidentification. Suppose I am facing a mirror and I believe that I see my own reflection when I really see someone else's. On the basis of what I see reflected in the mirror I say, "There's an incredibly tasteless painting hanging on the wall directly behind me." Because I believe that I am the person reflected in the mirror, I take 'me' to refer to the person reflected in the mirror. Thus there is a sense in which I take someone who is not myself, namely, the person reflected in the mirror, to have made my utterance and, hence, to be the referent of that use of 'me'. As is usual with cases of error through misidentification, we can interpret the utterance so as to make it come out true or false, depending on what referent we prefer to assign to the offending expression. Just as Henry Brown said either something true of Rebecca Thompson or something false of the dean, I have said either something true of the person reflected in the mirror or something false of me. But the parallel is imperfect. For there is almost no temptation to think that Brown actually intended to refer to the dean despite the fact that he used the expression 'the dean', and an almost overwhelming temptation, therefore, to interpret his utterance as true of Thompson. Not so with 'me'. We are inclined to say that I must have intended to refer to myself when I used the expression 'me' and that I thought that *in* referring to myself I was also referring to the person reflected in the mirror. Hence it is more natural to interpret my utterance as

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

false of myself rather than as true about the person reflected in the mirror. Nevertheless, the alternative interpretation might be warranted in certain circumstances, and, when it is, we have a use of the first person that incorporates error through misidentification.

We must grant, then, that, although uses of 'I' can be subject to referential error through misidentification, it is extremely rare that they are. There is always the strongest presumption that a speaker intends to speak of *herself* in using 'I'. With other referring terms, it seems, there is more scope for error through misidentification and, hence, more scope for supposing that a speaker intends to speak of something that is not the standard semantically determined referent of the term she uses.[9] Why is there so little scope for this in the case of the first person? Or, to put the question in a slightly different form, why is it that when I have false beliefs about who I am (as in the belief that I am the person reflected in the mirror), my use of 'I' still refers to me and not to the person of whom the beliefs are true? Is it because my capacity to refer to myself in the first person does not depend on my having true beliefs about myself in quite the way that my capacity to refer to other objects does depend on whether I have true beliefs about them?[10] If that were so, we would have a prima facie ground for supposing that first-person reference is indeed direct in a way that other reference is not. But there is an alternative explanation: that speakers generally have enough true beliefs about themselves in virtue of which they are quite clear about their identities even though they have some false beliefs about themselves as well. As I face the mirror, falsely believing that I am the person reflected in it, I know quite well who I am despite my false belief, and that knowledge surely comes into play when I use 'I' or 'me'. The kind of knowledge I mean here does not amount to some ineffable understanding that "I am me," but, rather, to genuinely descriptive knowledge about which thing in the world I am, knowledge about my past, my spatial location, my relations to others, etc.

All I have pressed so far is a conditional: if speakers could generally be relied upon to have an adequate body of true beliefs about

[9] But not too much scope so long as the principle of charity reigns. The point, however, still stands. We might say that the first person imposes a requirement of more perfect charity than other terms do. The reason why will emerge shortly.

[10] I speak in the strong terms of belief and knowledge rather than in the weak terms of capacities and dispositions because the necessary capacities and dispositions that in general make the individuation of objects possible (thereby performing the task of Frege's sense) cannot be described independently of the beliefs and knowledge to which they give rise. I therefore find it simpler to speak in directly epistemic terms. Throughout the paper my uses of phrases like 'background of beliefs', when used to characterize what mediates reference, can be harmlessly replaced by phrases like 'disposition to form beliefs'.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

themselves, that would explain why uses of 'I' are rarely hampered by referential error through misidentification. I have also suggested that this explanation is more plausible than one that invokes a notion of direct reference unmediated by beliefs or criteria for one's identity, since that would imply that 'I' is absolutely immune from error through misidentification, and we have seen that it is not. Does the antecedent of my conditional have independent plausibility? That is, can we give non-question-begging reasons for believing that speakers are in general clear about their identities? I think we can, though I admit that the claim is not obviously true. Two reasons for disbelieving it spring readily to mind. First, we must admit that amnesiacs and those who are seriously deluded are capable of using the first person, and they have relatively few true beliefs about who they are. Second, we can imagine purely inner uses of 'I' in epistemically impoverished settings such as the dark of Cartesian doubt (a more vivid setting is one that Anscombe envisages in which an amnesiac under sensory deprivation entertains thoughts in the first person). The second reservation requires separate treatment, and I will address it in the next section.

For now, let me give a reason for supposing that anyone capable of *public* uses of 'I' must in general have enough true beliefs about herself to avoid referential error through misidentification.

If a subject is to be capable of successful communication, she must generally know that she speaks when she speaks, and she must also generally know that another—an interpreter—could recognize that she speaks when she speaks. Since we are confining our discussion to public speech, we can safely affirm that the first bit of knowledge is not possible without the second; that is, to know that one is speaking is just to know that another could recognize that one is speaking. Insofar as a subject is to be capable of extended conversation, not only must she know that an isolated utterance of hers could be recognized as such by another, but also that consecutive utterances could all be correctly recognized as issuing from the same speaker—herself. The word 'correctly' is crucial here. Unless an interpreter could *correctly* reidentify the speaker, there could be no extended conversation between the interpreter and *her*. Thus we have, emerging as central to a speaker's capacity to communicate, an implicit understanding of some minimal and suitably public criteria for her identity over time, criteria by which an interpreter could correctly identify her. Moreover, this understanding enters into first-person reference. For, in knowing that she refers to herself when she uses 'I', a speaker knows she refers to "the speaker," and, in knowing she is "the speaker," she knows what an interpreter knows in know-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ing she is "the speaker." Viewed in this light, first-person reference looks strikingly indirect.

The foregoing provides something of an answer to Anscombe's query: What does a subject know when he knows that he refers to *himself* when he uses 'I'? The important point not to lose sight of is that it seems to be a condition on communication that speakers can be correctly individuated and reidentified by others (in the short term at least) and that they understand this. Even amnesiacs and the seriously deluded, if they are capable of a public use of the first person (if they speak and do not rave), know this much about "who they are." How does this explain why we rarely ascribe referential error through misidentification to speakers when they use 'I' (even on pain of rendering their utterances false)? It does because we have just established that, in any given speech situation, the speaker must have an adequate grasp of her identity for the purpose at hand. To call into doubt that a speaker is correctly applying an expression governed by the rule that it refers to the speaker is to call into doubt that the speaker can speak at all. Or, to put it the other way round, we cannot have a ground for interpreting an expression as governed by the rule that it refers to the speaker without *ipso facto* having a ground for supposing that the term is correctly applied.[11]

II

From certain philosophical perspectives (e.g., the Cartesian perspective), the discussion in section I overlooks what is truly important about and distinctive of the first person, and that is the way in which its use in inner life can be prised apart from any public criteria for one's identity. We might say, for example, that, antecedent to communication, a subject has an awareness of herself as a thinking thing which is quite independent of understanding how an interpreter may identify her, the latter being additional knowledge required only for public uses of 'I' and not for self-reference in thought. Indeed, my remarks at the end of section I appear, when regarded from this point of view, to lead to a hopelessly regressive account of first-person reference. I have explained a linguistic phenomenon—the immunity of public uses of 'I' from referential error through misidentification—by appeal to a background of true beliefs about oneself or one's identity (or a capacity to individuate oneself; see footnote 10). Should we not think of these beliefs (or capacities) as already presupposing a direct first-personal access to oneself in thought which underlies them? At this juncture we might well want to invoke

[11] It should be clear that this constraint on interpretation is stronger than the over-all constraint of charity. The latter requires that speakers emerge as having largely true beliefs. The former is stronger because it is more specific, saying that certain beliefs (about oneself) *have* to be among the true ones.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the idea of a direct and exclusively first-personal way of referring to oneself which is completely different from mediated reference to other things. If we did so, we would likely experience a strong temptation to see this fundamental reference to the self as mediating reference to all other things.[12] This of course would not be possible if first-person reference was itself mediated. If we took this course we would also need to invoke the idea of a special sort of object—the self, soul, ego, pure subject of consciousness—to serve as the referent.[13] Some philosophers deny that this inner use of 'I' is possible without a conception of oneself as an embodied person. Others take the possibility seriously, but find the idea of direct reference to a special object known only inwardly so philosophically unpalatable that they simply deny that this inner use refers at all. This is why in Kant's view the 'I' of the 'I think' does not stand for an object, why what Wittgenstein called the 'use-of-I-as-subject' is not a referring use, and why Anscombe, good disciple of Wittgenstein that she is, follows him in this.

It does not seem to me that the use of 'I' in inner life *can* be prised apart from one's conception of oneself as a public object. But for the sake of argument I want to take the possibility seriously, so as to show that it is not at all clear that we would have to posit a mysteriously direct way of referring to ourselves in order to account for it. We may, however, have to posit direct awareness or knowledge of our inner subjective states. One major difficulty with this idea is that a large part of our conception of such states is their role in bringing about and explaining our actions. It is implausible to suppose that subjects could have knowledge of such states without conceiving of themselves as agents, and that seems to imply a conception of themselves that goes beyond the purely inward knowledge afforded by

[12] This is precisely what Schiffer argues, *op. cit.* Chisholm and Lewis seem to follow him in this, though less clearly, since their concern is not reference but rather a uniform analysis of the psychological attitudes. Each proposes to reduce *de dicto* and *de re* attitudes to *de se* attitudes.

[13] If we did not posit a special object to serve as the referent of 'I', that is, if we insisted that 'I' refers to a public object (the person), the problem of explaining the supposed directness and priority of first-person reference would be insurmountable. For the identification of a person would require locating her in space, which in turn would require relating her to other (already identified) objects in space. See P. F. Strawson, *Individuals* (New York: Doubleday, 1959) and Gareth Evans, *Varieties of Reference* (New York: Oxford, 1982) for an account of the interdependence of self-identification and the identification of other bodies. (Evans's position is complex because it is at once Fregean and Russellian. Thus Evans insists on the above interdependence and also on the claim that the self-identification underlying uses of 'I' does not depend on explicit beliefs of the form "I am identical with. . . ." Nevertheless, the capacity to form and adjust such beliefs is required by the account, and so it goes some distance in meeting the Fregean requirement.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

their "use-of-I-as-subject."[14] But even if we suspend our doubts on this score, I want to emphasize Hume's point that direct awareness or knowledge of one's subjective states does not necessarily come to the same thing as direct acquaintance with their subject.

Consider a subject whose situation actually gives a point to Cartesian doubt, the sensorily deprived amnesiac who entertains a thought in the first person. She does not know whether she has a body or whether there is anything in the world besides herself. She does not even know what she is. If we are to make any sense at all of her entertaining first-person thoughts, we must suppose that, like Descartes, she is aware of her thoughts and, in being aware of them, is also aware that they must have a subject and that she must be that subject. She would then be in a position to refer inwardly to herself, the subject of her thoughts, with something like our pronoun 'I'. Philosophers of the modern period have tended to think that what lies behind such uses of 'I' is direct awareness of the *activity* of thinking and that this includes a direct awareness of *oneself* thinking. But we do not need such a model in order to account for the inner use of 'I'. We might give the term 'awareness' a weak reading according to which a subject, in being aware of thinking, is aware of herself thinking, though she may not explicitly separate out or form the idea of herself as distinct from her thoughts. Yet we can insist that this is something she must do if she is to refer to herself with 'I', rather than to her thoughts. How might she accomplish this? That depends on what we imagine her inner "awareness" to be like, and we can imagine it in either of two ways. First, her thoughts might be relatively unconnected, in which case her conception of herself as a subject of inner states might not be that of an abiding subject of diverse, successive thoughts. Or, second, she might be able to remember and connect her various subjective states so as to conceive of them as all belonging to a single subject.

If the first is the condition our amnesiac is in, it is difficult to see how she could conceive of herself as distinct from her thoughts, since the context provides no way of conceiving of her own identity conditions as distinct from the identity conditions of a thought. But let us suppose that she can. Then, each time she thinks, she can refer to herself as the subject of that thought. Thus, underlying each use of 'I' is the capacity to form a belief (but only one), the belief that she is

[14] There is another difficulty which deserves mention but really requires separate treatment, and it concerns the connection between the individuation of mental states and a grasp of the identity of their owner. Strawson and Evans both insist that mental states literally owe their identity to the identity of their owner and, hence, cannot be used to individuate their owner. Although I have some sympathy with this view, I think it needs a good deal of qualification, and so will not rely on it here.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the subject of the current thought. Because there is only one "way in which the object (herself) is reached by her" there is no room for referential error through misidentification. So here we have a use of 'I' that would (were it a genuine possibility) be absolutely immune from referential failure. Does it refer directly in a way that no other referring term can? We might call it direct reference because the referent is individuated in only one way; the amnesiac has no background of beliefs about the referent. But directness in that sense is not the exclusive property of the first person, nor does it run counter to Fregean instincts. Granted, this inner use of 'I' does depend on direct awareness of the referent's attributes (in some loose sense of 'attribute'); but that leaves room for the idea that the referent of 'I' is individuated in *terms* of those attributes, and hence for the view that the 'I' of inner life refers in precisely the way that other terms do (i.e., mediatedly).

What if our amnesiac's consciousness is connected by memory? (We can imagine that there is a definite point in the past, say, the moment sensory deprivation set in, beyond which she has no memories.) Surely, if she has a conception of herself as the abiding subject of successive thoughts and inner states, she is capable of judgments like the following:

> I am the subject of the present memory (at t_n), which is a memory of a certain thought (at t_m), and I am identical with the subject of that thought (at t_m).

In other words, she may have beliefs about her different "attributes" at different times, and hence more than one way of individuating the referent of 'I'. This ought to imply that her use of 'I' could be subject to referential error through misidentification. Kant thought we have beliefs of the sort we have credited our amnesiac with, but he also thought the claim "I am the single subject of all the states known to me through inner sense" to be analytically true. That might be so, but only if we confine the use of 'I' to present-tense self-ascriptions of subjective states, as we did in the preceding paragraph. Once we introduce memory and the possibility of past-tense self-ascriptions of subjective states we can no longer take this for granted. At least not if we take Shoemaker's notion of "quasi-memory" to be possible, that is, not if a subject can be psychologically continuous with a subject without being identical with that subject.[15] For then the judgment above, in which our amnesiac identifies herself as the subject of a certain past thought on the basis of memory, could be false without

[15] See Shoemaker's "Persons and Their Pasts," *American Philosophical Quarterly*, VII, 4 (October 1970): 269–285.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

her realizing it. (She could fail to realize that she is psychologically continuous with but not identical with that subject.) Because of this false "background belief," her use of 'I' in "I am the subject of the thought at t_m" is not guaranteed to refer to her rather than to the subject with whom she is psychologically continuous. So once again we have no compelling reason to suppose that the use of 'I' in inner life refers in a direct way (refers unmediated by background beliefs or capacities to form them) that is peculiar to the first person. It seems to be mediated by beliefs about identity in much the way all reference is. Of course, by hypothesis, the "use-of-I-as-subject" on the part of a sensorily deprived amnesiac would depend on a kind of knowledge of her states or attributes that is exclusively first-personal. But even on that hypothesis (which I think ought in any case to be regarded with suspicion) we can employ our ordinary (Fregean) model of reference in order to account for the use of 'I'.

III

Thus far I have concentrated mainly on the referential guarantees surrounding uses of 'I', and I have argued that we can explain them without appeal to a notion of direct reference peculiar to the first person. But there are other distinctive features of the first person that remain to be investigated, in particular, the irreducibility of 'I' to any other (co-referring) name or description or set of descriptions.

Since I grant the irreducibility, the question for me is, what does it amount to? It amounts to this: for any name or definite description *F* that refers to me, I may not know that I am *F* and yet I could still refer to myself with 'I'. So 'I' could not be equivalent in meaning to any such name or description. The point as it stands does not bring out a feature that is peculiar to 'I' alone. It might apply, for instance, to 'the shortest spy'. Since I don't (to my knowledge) know any spies and since I have absolutely no idea who the shortest spy might be, I can truthfully assert that for any name or description *F* that refers to the shortest spy—except of course 'the shortest spy'—I do not know that it refers to the shortest spy. Therefore no other referring expression in my idiolect is equivalent to 'the shortest spy'. A defender of the view that first-person reference is direct might draw a moral from this example that would support her view. She might say that the expression 'the shortest spy' is irreducible in my idiolect because I have only one way of identifying or individuating its referent, namely, as the shortest spy. Likewise, 'I' is irreducible because each of us has only one (distinctively first-personal) way of identifying or individuating its referent, namely, as *oneself*.[16]

[16] This seems to be Castañeda's proposal.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

If we want to undermine this argument for the direct-reference view of the first person, the only strategy available is the usual one: find a way of accounting for the distinctive features of the first person—in this case its irreducibility—without invoking the notion of direct reference. In light of the discussions in sections I and II we might begin by insisting that in any use of 'I' there must be some individuating belief that gives substance to the idea that the subject has referred to one thing and not another (the belief that I am me is not such a belief). In support of this claim I offer the following observation: it may well be that, for any name or description *F*, I may fail to know that I am *F*, but it does not follow that I can know which thing I refer to with 'I' without knowing that I am *F* for *some F* (where '*F*' is not 'I'). This is especially plausible if we think of public uses of 'I' where the speaker intends an interpreter to recognize her as the referent of 'I'. But then why can we not find some *F* for each use of 'I' such that that 'I' 's meaning in that particular use is reducible to *F*? The reason, I think, is as follows. When I choose to refer to myself with 'I', part of my intention is to get an interpreter to recognize that its referent is identical with the one who is using it: the speaker. This is what the first person is designed for and what it expresses. I cannot get this across by using any other referring expression; for no other referring expression is designed for that purpose. This way of seeing the matter does allow that there is a special sort of referential intention I can have in referring to myself which I cannot have in referring to anything that is not myself (and which no one else can have in referring to me), but it does not require that first-person reference can be achieved independently of my capacity to individuate myself as a public object in accordance with some criteria for the identity of persons.

The worry I raised for my view at the beginning of section II (156) over the possibly regressive aspect of my account crops up again here. The special referential intention that a speaker can have only with respect to himself is admittedly irreducible, which is to say, not expressible without the emphatic reflexive. It won't do, for instance, to try to render it as follows: there is a recognition on the part of the speaker that the term 'I' refers to the speaker (whoever that may be). For the speaker must know that it is *he* who speaks. My point, however, is that this recognition or knowledge, though it undoubtedly involves a sort of reflexive self-awareness, is not possible independently of individuating oneself as an object.

IV

John Perry suggests that the irreducibility of 'I' to any other co-referring expression has little to do with whether or how 'I' refers. As he sees it, 'I' is set apart from other modes of reference because we use

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

it in the expression of certain beliefs and other attitudes that play a distinctive psychological role, a role in motivating action. The connection with the irreducibility of 'I' comes out in the following example of Perry's (which I embellish slightly).[17] Lingens is lost in a library and believes himself (in the first-personal or *de se* sense) to be so lost. He is suffering from amnesia and he wants to leave the library. He believes, of a person named "Lingens," that Lingens is located in aisle 5, floor 6 of the Stanford Library, and further, he believes that Lingens could leave the library by making two rights, taking the elevator to floor 1, and then going straight out the door. But even though Lingens has a belief about Lingens (under the name 'Lingens') which, in conjunction with his desire to leave the library, ought to motivate him to make two rights, take the elevator, etc., he fails to act. What Lingens lacks is the first-person or *de se* belief that he himself is Lingens, the belief he would express by saying "Oh, *I'm* Lingens." The general implication is that the possession of beliefs about myself can neither motivate me nor explain my actions unless they either include or are closely tied to first-personal beliefs about myself.[18]

For some this might indicate that the "essential" and ineliminable indexical 'I' refers without mediation by descriptive knowledge. Perry does not draw this conclusion. He is content to point out that 'I' stands apart because it has this distinctive motivational role that other modes of reference to oneself don't have. For most, including Perry, this indicates that we cannot give an explanatory role to the psychological attitudes so long as we take propositions to be their sole objects. But, ineliminable though 'I' is, and as essential as non-propositional beliefs may be to agency, it remains true that we won't have a proper understanding of the distinctive role of the first person in bringing about action unless we see how it is connected with descriptive beliefs that can reliably tell us which thing in the world we are and why 'I' refers to that thing (so described).

In order to drive this last point home, I want first to spell out how David Lewis's account of *de se* attitudes would apply to the Lingens

[17] From Perry, "Frege on Demonstratives," *Philosophical Review*, LVI, 4 (October 1977): 474–497.

[18] Perry is concerned about whether indexicality can in general be handled within a Fregean framework, and this is unsurprising since, as was noted in fn 3, p. 150, other indexicals play a motivational role similar to 'I'. This raises the question whether 'I' can be given separate treatment such as is attempted here. But what I say about 'I' applies equally well to the other indexicals. 'Here' and 'now' cannot pick out places and times in a vacuum, however immediate our grip on the here and now seems to be. The terms make no sense independently of the framework(s) of space and time, i.e., independently of the identification of other places and times. The need for the individuating work of sense is even more obvious in the cases of 'this', 'that', etc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

example. It is instructive to think the example through in Lewis's terms because he analyzes *de se* attitudes as the direct self-ascription of properties, and his way of describing the latter obscures its connection with the concept of a person and the way in which persons are individuated. On one conception of this self-ascription of properties, personhood is not even required for it, as far as Lewis is concerned. His main reason for suspending the requirement is simplicity. He prefers an account in which the self-ascription of properties is ascription *simpliciter,* that is, unqualified and not indexed to place or time or possible world. Many of the properties we self-ascribe last for only a short while, much shorter than the temporal span of our existence. Lewis holds that in such cases *we* are not really the believing subjects. Rather, time-slices of us are, slices whose lives are exactly as long as the ascribed properties last. Thus, in the case of a person who is wondering what the time is, it is really a very small time-slice of her which is wondering when it exists. It wants to be able to self-ascribe *simpliciter* the property of existing at a certain time.

Lingens can self-ascribe the property of inhabiting a world in which Lingens is lost in the Stanford Library and in which Lingens could leave the library by making two rights, etc. Lingens needs to self-ascribe the property of being Lingens, and, given what he knows about Lingens, he could self-ascribe that property if he could self-ascribe the property of being in aisle 5, floor 6 of the Stanford Library. All he is in a position to do is to self-ascribe the property of being in a certain perceptual situation. He needs to learn that that perceptual situation is the situation of being in aisle 5, floor 6 of the Stanford library. Now the property of being in a certain perceptual situation is not a property that Lingens can self-ascribe *simpliciter.* It must therefore be a time-slice of Lingens that self-ascribes that property. Perhaps, then, it is the time-slice that wants to leave the library, since the desired property is not one Lingens could have *simpliciter* either. On the other hand maybe it *is* Lingens who wants to leave the library, but the desire is really a desire to contain a certain (dated) time-slice which is outside the library. There are many more or less equivalent ways of construing the Lingens case in terms of time-slices. What all of them reveal is that Lingens's realization that he himself is Lingens is a complex affair. Moreover, they reveal that, ultimately, time-slices must identify themselves or individuate themselves by their relations to the person Lingens if their *de se* attitudes are to play the distinctive role generally ascribed to *de se* attitudes. First I will show how this emerges on the hypothesis that the majority (and perhaps all) of Lingens's beliefs and attitudes are really held by time-slices of Lingens, and that not Lingens but time-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

slices of Lingens are really the agents whose behavior we want to explain by appeal to those attitudes. Second, I will consider how the analysis would work if Lingens himself were reinstated as the agent.

On the first hypothesis we have a time-slice of Lingens, Little Lingens I, which self-ascribes the properties of being lost in a library and of living in a world where a person named Lingens has a time-slice lost in the Stanford Library. Little Lingens I cannot want the property of being outside the library unless it is irrational, for it knows it will not last that long. But Little Lingens I can want the property of being succeeded (at some fairly specific time in the near future) by a time-slice that is outside the library. Now what does Little Lingens I need to learn in order to satisfy this want? That it is located in the Stanford Library. Given what else Little Lingens I believes, this would amount to learning that it belongs to *Lingens*. If we adopt this analysis, we will want to add to our account of time-slices so as to ensure that there is some sort of attitudinal (both cognitive and conative) continuity between Little Lingens I and Little Lingens II. For, in general, what earlier time-slices of Lingens learn about themselves and Lingens must be passed on to the succeeding time-slices. And, for the most part, Lingens's time-slices will not want much for themselves. Rather, they will be especially concerned about the well-being of certain future beings, namely, future time-slices which also belong to Lingens. In short, Lingens's time-slices must be such that in aggregate they make up the unit Lingens is, and they must see themselves in that light, i.e., as having a role in a life that extends beyond their own. In other words, the ultimate reference point of their *de se* attitudes would not, strictly speaking, be themselves.

We can get the same result by turning the analysis around, though here much of the point of introducing time-slices is lost. Let Lingens and not his temporal parts be the agent who desires and acts on his desires. Most of Lingens's desires will concern his temporal parts. Lingens might want each of his temporal parts to be happy or he might sacrifice the happiness of some for the happiness of others. Much of the information Lingens needs in order to satisfy his wants will be information about his time-slices. If Little Lingens I can self-ascribe the property of being in Stanford Library or the property of belonging to Lingens, it will not do Lingens any good unless Lingens can know of Little Lingens I what Little Lingens I knows of itself. Only then can Lingens ensure the happiness of Little Lingens II by bringing it about that it is outside the library. In general, for every relevant property that a time-slice of Lingens self-ascribes, there will have to be a property that Lingens self-ascribes, namely, that of having a (dated) time-slice with that property.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

However we tell the time-slice story, the concept of a person in general and also the capacity to identify the particular person Lingens must be allowed a place. For every story must accommodate the intuitive idea that, when Lingens learns that he himself is Lingens, he learns which person he is. And every story must also accommodate the obvious truth that rational action presupposes psychological connectedness. Any behavior that deserves to be called an action must spring not just from one isolated psychological state but from several interrelated states, a combination of desires, beliefs, and other attitudes. The vehicles of this integration in Lingens's case are beliefs about the identity of the person Lingens. Ordinarily, we would say that it is because Lingens knows *de se* of himself that he is a *person* in a *library,* that he is capable of the realization that the person *Lingens* similarly lost in *Stanford* library is *himself,* where that realization can affect his actions. Endow a time-slice with as much capacity for *de se* attitudes as you wish. Unless it knows something about the person to whom it belongs, beliefs about the person Lingens can never be brought to bear on its situation. That would mean that nothing could play for it the distinctive role we attributed to Lingens's first-person belief that he himself is Lingens. In other words, it is a mistake to suppose that this role is possible without a rich (thicker-than-a-time-slice) conception of the self. Moreover, the fewer true *de se* beliefs Lingens (or his time-slice) has about what sort of thing he (Lingens) is and which thing he is, the less we can make of the idea that his *de re* beliefs about someone who is in fact himself can be replaced by the appropriate *de se* beliefs. And the less scope there is for such a conversion, the less scope there is for rational action.

This brings me to a more general point about the connection between having true beliefs about oneself and one's status as a rational agent. It is a commonplace that an agent's capacity for successful action depends on the possession (or at least the availability) of relevant knowledge, i.e., *true* beliefs. Lying at the heart of such knowledge must be true beliefs about which thing one is: where one is and what one's relations to various objects are. If an agent were entirely lacking in true beliefs of this sort even the simplest actions (such as movements) would be prone to fail, and her very status as an agent could be called into question. In that case the main (if not the whole) point of belief/desire explanation would be lost, and, more particularly, there would be no place for the *de se* attitudes that seem so essential to motivation.

V. CONCLUSION

Perry's explanation of the irreducibility of *de se* attitudes to propositional attitudes in terms of their distinctive psychological role is fine

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

as far as it goes. But it is misguided to suppose that this role can be prised apart from the sort of self-knowledge in virtue of which agents achieve their status as agents. We can now see that the first person is bound up in a deep way with individuating beliefs adequate to secure reference to oneself.

The last section provides a broader perspective from which to view the immunity (or at least resistance) of 'I' to referential error through misidentification. The question I raised in section I is this: Why, even when a subject has some false beliefs about herself, do we take her uses of 'I' to refer to her rather than to the thing of which her beliefs are true? In answering the question I appealed to very special features of communication. I suggested that communicators must have true identifying beliefs about themselves of a sort that makes it possible for them to fulfill their intentions to be interpreted *by others*. This turns out to be an application of a more general point about agents. Agents must in general have adequate true beliefs about their identity, and the respects in which they must correctly identify themselves depends on their purposes, the kinds of activity in which they engage. Communicators need knowledge about themselves adequate for their communicative purposes, and this requires knowing more than which spatiotemporal particulars they are. But the special (social) way in which communicators individuate themselves is in part supported by more fundamental beliefs about themselves in virtue of which they are agents in a broader sense, and these are perhaps by themselves adequate for fixing the referent of 'I'.

It might be suggested that I have overstated my case and that we need not be agents in order to self-refer: In other words, it could be argued (and it has been argued) that although the success of our actions depends on our having adequate true beliefs about ourselves, our capacity to self-refer does not so depend. For example, a sensorily deprived amnesiac and, also, a brain-in-a-vat are both merely potential agents, and both lack true beliefs of the sort I claim generally underlie first-person reference. But if they self-refer, it must be solely in virtue of their capacity for the "use-of-I-as-subject."[19] And

[19] The case of a brain-in-a-vat is quite different from the case of a sensorily deprived amnesiac, for a brain-in-a-vat may have a functional makeup analogous to (or equivalent to) the functional makeup of an embodied agent, including states whose functional role is analogous to the role of an agent's *de se* attitudes. On my view, we should regard these states of a brain-in-a-vat as self-referring only if we regard her other (non-first-personal) states as referring. But it is natural to think of a brain-in-a-vat as only potentially referring (as having intentional states that would be beliefs about objects if the brain were put into the right sort of context). I suggest that we view the brain's capacity for self-reference likewise—as merely potential. Only if we wanted to allow that the "use-of-I-as-subject" need not be bound up with

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

as I stressed in section II, we need not abandon a quasi-Fregean conception of reference (as requiring individuating beliefs) in order to account for the referring status of purely inner uses of 'I', nor should we suppose that we have to abandon that conception in order to account for the distinctive features of the first person even when we confine ourselves to the "use-of-I-as-subject."

My further point, however, is that concentrating on the use of 'I' in inner life leads us to overlook the way in which the conditions on rational agency and communication provide for the distinctive features of the first person as they manifest themselves in ordinary, public contexts. Each of them—the immunity of 'I' from referential error through misidentification, the irreducibility of 'I' to any name or description (or set of descriptions), and the distinctive motivational role of first-person attitudes—is more interesting when viewed in connection with agency and communication. For then we see clearly that the first person is bound up in a straightforward way with a background of self-individuating beliefs so as to satisfy in spirit, if not in letter, Anscombe's Fregean desideratum. All of this is more plausible if we suppose that only social, communicative creatures are capable of the sort of self-consciousness that 'I' expresses. For then, rather than view self-consciousness as an immediate or direct awareness of oneself independent of criteria for one's identity, we would see self-consciousness as simply an awareness of what *kind* of thing one is, i.e., a rational, communicative agent. What sets 'I' apart is not any direct way in which it refers, but rather, what it expresses about a subject's status as agent or communicator. A public use of 'I' expresses a kind of referential intention that a speaker can have only with respect to herself, and a private use of 'I' expresses one's disposition to act on one's belief; we might say it affords the connection between one's desire, one's beliefs about how a certain individual could satisfy that desire by performing a certain action, and oneself performing that action. Neither of these functions of the first person would be possible unless subjects were relatively clear about their identities. It is in this sense that it is constitutive of self-conscious subjects (subjects who can use the first person) that they have enough true beliefs about themselves to secure reference to themselves.

CAROL A. ROVANE

Vassar College

---

a "use-of-I-as-object" would we want to say that a brain-in-a-vat could refer to herself with 'I'. This is why I treat the case on a par with the case of a sensorily deprived amnesiac.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

A NOTE ON EXTENSION, INTENSION, AND TRUTH*

IT is common knowledge that two predicates may coincide in extension but differ in intension and that, for any predicate, one can construct an infinity of coextensional predicates that differ in intension. Hence the concept that a predicate expresses—and its intension—cannot be determined from the extension of the predicate alone. Something stronger also appears to be the case. If a predicate has the same extension as a given concept, then one cannot determine whether the predicate expresses the concept simply on the basis of extensional information. The purpose of this note is to show that this appearance is illusory. We will construct an example in which a predicate G belonging to a first-order language L has the same extension as a concept $\mathcal{C}$, and yet merely by knowing the extensions of the nonlogical constants of L one can determine that G does not express $\mathcal{C}$.

The example is based on the fact that under certain conditions—conditions under which certain kinds of vicious reference are absent—a classical language L can contain a predicate that is extensionally equivalent to its own truth. Such a predicate we shall call a *T-predicate for L*.

Here is one simple way of constructing a classical language that contains its own T-predicate. Begin with an ordinary first-order language with the usual logical resources ($=$, $\sim$, $\supset$, $\forall$, etc.), and add to it a one-place predicate G and quotational names for all the sentences. Let the resulting language be L. Take any model M of the G-free fragment of L that meets the following conditions: (i) the domain D of M contains all the sentences of L; (ii) the quotational names receive their intended interpretation; (iii) the interpretations of the predicates and the nonquotational names do not distinguish among the sentences of L.[1] This last condition means that no names other than the quotational names pick out a sentence of L; that if a one-place predicate is true of a sentence of L, it is true of every sentence of L; and similarly for the n-place predicates. Now, if we assign to G an interpretation, say g, we obtain a standard classical model $M + g$

* We wish to thank Mark Wilson and Michael Kremer for their valuable comments on the first draft of this note. Kremer helped clarify the role of symmetry in the argument and suggested the appropriateness of the term 'supervenience'.

[1] What we are calling "models" are sometimes called "structures." The interpretation of a predicate in a model M determines and is determined by its extension in M. For concreteness, we suppose below that the interpretation of a one-place predicate in M is a function from the domain of M into the truth values *The True*, *The False*.

0022-362X/87/8403/0168$00.70 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

for the whole of L. Let L_{M+g} be the resulting *interpreted* language. We can determine by the usual Tarskian method the sentences of L_{M+g} that are true and those which are false. Define an operation τ_M on the possible interpretations of G in M in the following way: for any interpretation g, $\tau_M(g)$ is the interpretation that assigns the value *The True* to the sentences true in L_{M+g} and the value *The False* to the rest of the objects in D. Observe that *G is a T-predicate for L_{M+g} if and only if g is a fixed point of τ_M* [i.e., if and only if $g = \tau_M(g)$]. It can be shown that τ_M has a fixed point and indeed a unique one.[2] If g is this fixed point then L_{M+g} is a classical language that contains a T-predicate for itself.

The language L_{M+g} constructed above has no vicious reference of any kind. Presence of certain kinds of vicious reference destroys the possibility of the above construction. For example, suppose that the model M' is just like M except that in M' an individual constant b denotes the sentence $\sim Gb$. It can be shown that $\tau_{M'}$ has no fixed points. (The argument here parallels the proof of Tarski's indefinability theorem.) So there is no interpretation g that makes G a T-predicate for $L_{M'+g}$. On the other hand, presence of certain other kinds of vicious reference leaves intact the construction given above. It is this that we exploit in our example.

Suppose that L contains, in addition to the other constants, a name c. Suppose that M^* is a model of the G-free fragment of L that meets the conditions (i) and (ii) given above and, apart from the constant c, meets condition (iii) also. Finally, suppose that the denotation of c in M^* is Gc. Now one can show that τ_{M^*} has exactly two fixed points. With one of them, say g, we obtain a classical language L_{M^*+g} in which Gc is true. With the other, say g', we obtain the classical language $L_{M^*+g'}$ in which Gc is false.

Imagine a community C in which the language L is so used that its interpretation is, as a matter of fact, given by $M^* + g$. Imagine, in particular, that the predicate G in the language of this community means "is a sentence that the computer XYZ is programmed to

[2] We have assumed for simplicity that the language contains no function symbols. This assumption is dispensable at the cost of adding a fourth condition to the three stated above. Further, the conditions given can be considerably weakened. For more details on this and for a proof of the existence and uniqueness of the fixed point, consult Anil Gupta, "Truth and Paradox," *Journal of Philosophical Logic*, XI, 1 (February 1982): 1–60; also reprinted in Robert L. Martin, ed., *Recent Essays on Truth and the Liar Paradox* (New York: Oxford, 1984). The idea of fixed points, in the context of a three-valued approach to the theory of truth, was introduced most notably by Robert L. Martin and Peter Woodruff, "On Representing 'True-in-L' in L," *Philosophia*, V, 3 (July 1975): 213–217; and by Saul Kripke, "Outline of a Theory of Truth", this JOURNAL, LXXII, 19 (Nov. 6, 1975): 690–716. Both of these papers are reprinted in Martin's *Recent Essays*, *op. cit.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

produce," and that, as a matter of fact, the computer is programmed to produce all (and only) the true sentences of L_{M^*+g}, including the sentence Gc. Since g is a fixed point of τ_{M^*}, the interpretation of G in this community's language is given by g. So the predicate G as used by this community is coextensional with the following two (intensionally nonequivalent) concepts:

being a true sentence in the community's language ($\mathcal{C}$)
being a sentence that the computer XYZ is programmed to produce ($\mathcal{C}'$)

We will argue that G cannot mean $\mathcal{C}$. A linguist who discovers that the predicate G in the community's language is coextensional with the concepts $\mathcal{C}$ and $\mathcal{C}'$ will have reasons to conclude, on the basis of the extension of G alone, that it does not express $\mathcal{C}$. The following terminology will prove useful: let us say that a predicate of L is a *truth-predicate* if and only if it means "true in L." Our claim, then, is that, although G is a T-predicate for the community's language, it cannot be its truth-predicate.

Let us observe first that the hypothesis that G expresses $\mathcal{C}'$ has certain virtues over the hypothesis that G expresses $\mathcal{C}$. On the first hypothesis Gc says of itself that it is a sentence that the computer XYZ is programmed to produce, and it is true because the computer *is* programmed to produce Gc. There is no mystery why Gc is true. On the second hypothesis Gc says of itself that it is true and, thus, is like the Truth Teller in English: "This very sentence is true." Now it is a mystery why Gc is true in the community's language. We can—in some sense of 'can'—say that the Truth Teller is true, and we can also say that it is not true. But there is little to choose between these options. Both are equally arbitrary. So, on the hypothesis that G expresses truth, we have the puzzle: how come the Truth Teller is true and not false in the community's language? This observation motivates, we hope, our claim. It does not prove the claim. However, the hypothesis that G expresses $\mathcal{C}$ conflicts with a basic feature of truth. We call this feature *the supervenience of the semantic interpretation of truth*. It is related to, but distinct from, Ramsey's Redundancy Theory.

It has often been remarked that the sentence " 'snow is white' is true" says nothing more nor less than the sentence 'snow is white'; that 'is true' is eliminable from and redundant in the first sentence. Whether this observation can be extended to all the sentences of a language containing its own truth-predicate depends on the logical resources of the language. If the language has propositional variables and quantifiers, the redundancy thesis is probably true. But if such quantification is not available and the language has ordinary

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

first-order quantifiers, the redundancy claim cannot hold for all the occurrences of 'true'. Nevertheless, even here the following kind of redundancy does obtain: *to evaluate the sentences of a first-order language that contains its own truth-predicate one does not need the semantic interpretation of the truth-predicate.* For example, suppose we wish to determine the semantic status of a sentence such as

(1) $(\forall x)(Fx \supset Gx)$

in a model. Suppose we are given the domain of the model and the interpretation of F. Suppose also that F is true of just one object in the domain: the sentence Hd. Now, to determine the semantic status of (1), we would need to know the interpretation of G in the model. This is so *except* when G is a truth-predicate. In this case the status of (1) is entirely determined by the interpretations of F, H, and d. More generally, the point is that, if a first-order language contains a truth-predicate G, then the status of all the sentences of the language is entirely determined by the semantic interpretation of the G-free part. The interpretation of G—whatever it be: two-valued or three-valued or via a rule of revision or some other[3]—supervenes upon the interpretation of the G-free part. This feature we are calling "the supervenience of the semantic interpretation of truth".[4] In reference to the community C imagined a little earlier, this feature implies that if G is a truth-predicate then the interpretation of G is fixed by the interpretation of the G-free part, i.e., by the model M^*.

Let us now imagine the community C in a slightly different situation s'. (Let us call the situation considered earlier s.) Suppose that in s' the meanings of the various constants in C's language are as in s; in particular, G in s' still means "is a sentence that the computer XYZ is programmed to produce." Suppose that, in other respects also, s' is just like s except that in s' the computer XYZ is programmed to

[3] Two-valued theories have been sketched by Charles Parsons and Tyler Burge; three-valued approaches were developed, as already indicated, by Kripke, Martin and Woodruff, and others; the revision approach was initiated by Gupta, Hans Herzberger, and Belnap. For these and other approaches to the theory of truth consult the papers reprinted in Martin's *Recent Essays*. The point made in this note holds, as far as we can tell, for all approaches.

[4] The supervenience described above is a special case of a much more general feature of truth. Suppose that $L_1, \ldots, L_n$ are languages that contain truth-predicates for each other. (One extreme possibility here is that they form the bottom part of a Tarski hierarchy. At another extreme each may contain truth-predicates for all the others.) Now, to determine the semantic status of the sentences of $L_1, \ldots, L_n$, one does not need the interpretation of the truth-predicates. These are redundant once the interpretations of the other constants of $L_1, \ldots, L_n$ are fixed.

Note that supervenience does not hold, in general, of T-predicates, but only of truth-predicates.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

produce the sentence $\sim Gc$ and all the rest of the truths of $L_{M^*+g'}$. As before, since g' is a fixed point of τ_{M^*}, G is coextensional with the concepts $\mathcal{C}$ and $\mathcal{C}'$. Now observe that the hypothesis that G is a truth-predicate in s' is no less (or more) plausible than the hypothesis that it is a truth-predicate in s. The two cases are exactly parallel. The only difference between them is that with the first hypothesis one is led to say that the Truth Teller is true and, with the other, that the Truth Teller is false. Neither of these consequences is any worse than the other. However, the supervenience feature discussed above implies that one cannot accept both of the hypotheses. For the G-free part of the language has the same interpretation in the two situations: its interpretation is fixed by M^*. So our hypothetical linguist would know, simply on the basis of the extension of G, that both of the hypotheses cannot be true. Symmetry of the two hypotheses leads to the conclusion that in neither situation is G a truth-predicate.

There are two features of the above argument that will be of concern even to a sympathetic reader. First, the claim of symmetry of the two hypotheses will appear debatable. In fact, some philosophers maintain that the Truth Teller is simply false. (The view that it is simply true is not found in the literature, as far as we know.) These philosophers will not regard the two hypotheses as equally plausible. They would favor the latter hypothesis over the former. Secondly, in the argument given we have not carefully distinguished between what is the case in a situation from what the hypothetical linguist knows to be the case. It may appear that the linguist needs to know more than we attribute to him, if he is to arrive at the desired conclusion.

Concerning the first point: Notice that our original claim—that the linguist will be able to conclude in the first situation s that G is not a truth-predicate—must be granted by anyone who denies symmetry and holds that the Truth Teller is false. Supervenience now implies that, if the interpretation of the G-free part of the language is given by M^* and G is a truth-predicate, then the interpretation of G must be g'. So the linguist will be able to conclude that although G is a T-predicate, it is not a truth-predicate. For the interpretation of G in s is g, and, had G been a truth-predicate, its interpretation would have been g'.

By similar reasoning, if the Truth Teller is held true, then s' yields the desired example. Hence, even if symmetry be denied, it must be admitted that, in one of s and s', G agrees in extension with truth although it is not a truth-predicate. The role of symmetry in our argument is minimal. It entitles us to say that, in *both* s and s', G fails

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to be a truth-predicate. It is, however, not needed in order to show the existence of the desired example.

Concerning the second point: We need to reconstruct the argument from the viewpoint of the linguist—call her Sally. Recall that Sally knows the extension of G in s and she knows that G is coextensional with the concept $\mathcal{C}$. This enables her to conclude that the interpretation (in s) of the G-free part of the language is given by a model M^* (she may not know which one) and that τ_{M^*} has at least one fixed point (for she knows that G is coextensional with the truth-concept $\mathcal{C}$). Since the language has quotational names and identity, she knows by examining the extension of $\mathcal{C}$ that Gc is self-referential and that this is the only kind of vicious reference in the language. [She needs to know that the quotational names have their standard interpretation and that the predicates meet condition (iii) stated above. Since this does not require any intensional information about the language, we may assume that she does know this.] So she concludes that τ_{M^*} has exactly two fixed points g, g', and that one of these, say g, gives the actual interpretation of G. (We do not need to assume that she knows anything about these fixed points other than that Gc is true in one and false in the other.) Believing in symmetry, she argues that, if it is possible that G be a truth-predicate in the community's language, then it is equally possible that G be a truth-predicate when the interpretation of the language is given by $M^* + g'$. So the hypothesis that G *may* be a truth-predicate in C's language, she reasons, yields the existence of worlds w and w' such that in w (w') the interpretation of the whole of the language is given by $M^* + g$ ($M^* + g'$) and in both of the worlds G is a truth-predicate. But this violates the supervenience of truth. She concludes that G cannot be a truth-predicate. At the cost of greater tedium, this argument can be made more rigorous and precise. Let us hope the reader is satisfied with the above.

When a language contains its own truth-predicate, the presence of self-reference and cross-reference can result in pathological sentences such as the Truth Teller and the Liar ("This very sentence is not true") whose semantic status is clear neither to intuition nor to theory. It should be observed, however, that the argument given relies on no particular judgment on these troublesome sentences. The only property used is one that both intuition and theory agree on: the supervenience of the semantic interpretation of truth.

That a T-predicate for a language should sometimes fail to be its truth-predicate is not in the least surprising. For "T-predicate" is an extensional notion, whereas "truth-predicate" is an intensional one.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Any predicate that is coextensional with truth qualifies as a T-predicate, but, to be a truth-predicate, stronger conditions have to be met. What is curious about the examples given above is that sometimes one knows on the basis of extensions alone that a T-predicate does not express truth. In these examples a T-predicate is coextensional with truth only because it does not mean truth. If it *had* meant truth, its extension would *not* have been what it actually is.

Even within an extensional perspective, "T-predicate" and "truth-predicate" invoke different procedures, different points of view. To determine whether G is a T-predicate, we take the interpretation of G *as given* and using that interpretation we see whether G is coextensional with truth. If it is, then G is a T-predicate; otherwise it is not. But when we entertain the hypothesis that G is a truth-predicate, no interpretation of G is taken as given (because of the supervenience feature). We determine on the basis of the G-free part alone, the sentences that are true and those which are not true. Sometimes the two ways of proceeding give the same result. Sometimes, as in the examples constructed above, they give different results. It is this that enables us to infer from extensions alone that G is not a truth-predicate.

Examples of the sort given above can be constructed using other kinds of vicious reference, other kinds of logics, and other semantic concepts. In particular, examples of this sort can be given using the three-valued fixed points investigated by Saul Kripke, Robert Martin and Peter Woodruff, and others.

ANIL GUPTA

University of Illinois/Chicago

NUEL BELNAP

University of Pittsburgh

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NEW BOOKS

NEW EDITIONS AND TRANSLATIONS

AVERROES: *Averroes' De Substantia Orbis.* Critical Edition of the Hebrew Text, with English Translation and Commentary. ARTHUR HYMAN, trans. & ed. Cambridge, Mass.: Medieval Academy of America; Jerusalem: Israel Academy of Sciences and Humanities, 1986. English, 156 p.; Hebrew, 74 p. $30.

BLOCH, ERNST: *Natural Law and Human Dignity.* DENNIS J. SCHMIDT, trans. Cambridge, Mass.: MIT Press, 1986. xxx, 323 p. $25.

BLUMENBERG, HANS: *Work on Myth.* ROBERT M. WALLACE, trans. Cambridge, Mass.: MIT Press, 1986. xi, 685 p. $40.

BOUWSMA, O. K.: *Wittgenstein Conversations 1949–1951.* J. L. CRAFT and RONALD E. HUSTWIT, eds. Indianapolis: Hackett, 1986. xxxiv, 78 p. Cloth $15, paper $4.95.

EMERSON, RALPH WALDO: *Emerson on Transcendentalism.* EDWARD L. ERICSON, ed. New York: Ungar, 1986. xiii, 109 p. $5.95.

FISCH, MAX H.: *Peirce, Semeiotic, and Pragmatism.* KENNETH LAINE KETNER and CHRISTIAN J. W. KLOESEL, eds. Bloomington: Indiana UP, 1986. xi, 464 p. $45.

GADAMER, HANS-GEORG: *The Idea of the Good in Platonic-Aristotelian Philosophy.* P. CHRISTOPHER SMITH, trans. New Haven, Conn.: Yale, 1986. xxxi, 182 p. $20.

GARDIES, JEAN-LOUIS: *Rational Grammar.* KEVIN MULLIGAN, trans. Washington, D.C.: Catholic University of America Press, 1986. 339 p. $59.95.

GÖDEL, KURT: *Collected Works, Volume I. Publications 1929–1936.* SOLOMON FEFERMAN *et al.*, eds. New York: Oxford, 1986. xvi, 474 p. $35.

GREEN, T. H.: *Lectures on the Principles of Political Obligation and Other Writings.* PAUL HARRIS and JOHN MORROW, eds. New York: Cambridge, 1986. ix, 383 p. Cloth $44.50, paper $14.95.

GUEROULT, MARTIAL: *Descartes' Philosophy Interpreted according to the Order of Reasons.* ROGER ARIEW, trans. Minneapolis: Minnesota UP, 1985. Vol. I: *The Soul and God,* xii, 327 p. Cloth $35, paper $15.95. Vol. II: *The Soul and the Body,* 311 p. Cloth $35, paper $15.95.

HANSLICK, EDUARD: *On the Musically Beautiful: A Contribution towards the Revision of the Aesthetics of Music.* GEOFFREY PAYZANT, trans. & ed. Indianapolis: Hackett, 1986. xxiv, 127 p. Cloth $16.50, paper $5.95.

HEGEL, G. W. F.: *Hegel Lectures on the Philosophy of Religion.* PETER C. HODGSON, ed. Vol. III: *The Consummate Religion.* Berkeley: California UP, 1985. xvii, 423 p. $38.50.

HEGEL, G. W. F.: *The Philosophical Propaedeutic.* A. V. MILLER, trans. MICHAEL GEORGE and ANDREW VINCENT, eds. New York: Basil Blackwell, 1986. xxxv, 175 p. $45.

JAMES, WILLIAM: *William James: Selected Unpublished Correspondence, 1885–1910.* FREDERICK J. DOWN SCOTT, ed. Columbus: Ohio State UP, 1986. x, 603 p. $45.

JASPERS, KARL: *Basic Philosophical Writings, Selections.* EDITH EHRLICH, LEONARD EHRLICH, and GEORGE B. PEPPER, eds. Athens: Ohio UP, 1986. xii, 556 p. Cloth $55, paper $26.95.

Digitized by Arya Samaj Foundation Chennai and eGangotri

KANT, IMMANUEL: *Kant's Latin Writings, Translations, Commentaries, and Notes.* LEWIS WHITE BECK, trans., with MARY J. GREGOR, RALF MEERBOTE, JOHN A. REUSCHER. New York: Peter Lang, 1986. 251 p. $39.

LEFORT, CLAUDE: *The Political Forms of Modern Society: Bureaucracy, Democracy, Totalitarianism.* Cambridge, Mass.: MIT Press, 1986. 342 p. Cloth $35, paper $14.50.

LEKTORSKY, V. A.: *Subject, Object, Cognition.* S. SYROVATKIN, trans. Moscow: Progress Publishers, 1984. 280 p. $7.95.

MARTIN, GOTTFRIED: *Arithmetic and Combinatorics: Kant and His Contemporaries.* JUDY WUBNIG, trans. Carbondale: Southern Illinois UP, 1985. xiv, 198 p. $27.50.

MARX, KARL: *Karl Marx: A Reader.* JON ELSTER, ed. New York: Cambridge, 1986. vi, 345 p. Cloth $34.50, paper $9.95.

NIEBUHR, REINHOLD: *Leaves from the Notebook of a Tamed Cynic.* San Francisco: Harper & Row, 1986. viii, 152 p. $7.95.

NIETZSCHE, FRIEDRICH: *Human, All Too Human.* R. J. HOLLINGDALE, trans. New York: Cambridge, 1986. xix, 395 p. $10.95.

RICOEUR, PAUL: *Time and Narrative,* vol. II. KATHLEEN MCLAUGHLIN and DAVID PELLAUER, trans. Chicago: University Press, 1985. vii, 208 p. $22.50.

RICOEUR, PAUL: *Lectures on Ideology and Utopia.* GEORGE H. TAYLOR, ed. New York: Columbia, 1986. xxxvi, 353 p. $35.

RUSSELL, BERTRAND: *The Collected Papers of Bertrand Russell,* vol. VIII. JOHN G. SLATER, ed. London: Allen & Unwin, 1986. xl, 418 p. $60.

SCHMITT, CARL: *Political Romanticism.* GUY OAKES, trans. Cambridge, Mass.: MIT Press, 1986. xxxv, 177 p. $20.

ANNUALS AND PROCEEDINGS

ASQUITH, PETER D., and PHILIP KITCHER, eds.: *PSA 1984, Volume Two.* East Lansing, Mich.: Philosophy of Science Association, 1985. xxii, 902 p. $30.

CLEARY, JOHN J., ed.: *Proceedings of the Boston Area Colloquium in Ancient Philosophy.* Vol. I, 1985. Lanham, Md.: University Press of America, 1986. xlviii, 330 p. $16.25.

FRENCH, PETER A., THEODORE E. UEHLING, JR., and HOWARD K. WETTSTEIN, eds.: *Midwest Studies in Philosophy X: Studies in the Philosophy of Mind.* Minneapolis: Minnesota UP, 1986. 680 p. Cloth $45, paper $18.95.

GIANCOTTI, EMILIA, ed.: *Proceedings of the First Italian International Congress on Spinoza.* Naples: Bibliopolis, 1985. xii, 537 p. $60.

HALPERN, JOSEPH Y., ed.: *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the 1986 Conference.* Los Altos, Calif.: Morgan Kaufman, 1986. vii, 407 p. $18.95.

HUMBER, JAMES M., and ROBERT F. ALMEDER, eds.: *Biomedical Ethics Reviews 1985.* Clifton, N.J.: Humana Press, 1986. xii, 184 p. $29.50.

MCMURRIN, STERLING M., ed.: *The Tanner Lectures on Human Values,* vol. VII. Salt Lake City: Utah UP, 1986. 208 p. $20.

TYMIENIECKA, ANNA-TERESA, ed.: *Poetics of the Elements in the Human Condition: the Sea.* Boston: Reidel, 1985. xv, 508 p. $69.

TYMIENIECKA, ANNA-TERESA, ed.: *The Phenomenology of Man and of the Human Condition,* part II. Boston: D. Reidel, 1986. xi, 401 p. $87.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ANTHOLOGIES IN PAPERBACK

ALLISON, DAVID B., ed.: *The New Nietzsche.* Cambridge, Mass.: MIT Press, 1985. xxviii, 274 p. $7.95.

ARTHUR, JOHN, ed.: *Morality and Moral Controversies.* Englewood Cliffs, N.J.: Prentice-Hall, 1986. viii, 470 p. $20.95.

BAR-ON, A. ZVIE, ed.: *On Shmuel Hugo Bergman's Philosophy.* Amsterdam: Rodopi, 1986. 134 p. $19.95.

BAYNES, KENNETH, JAMES BOHMAN, and THOMAS MCCARTHY, eds.: *After Philosophy.* Cambridge, Mass.: MIT Press, 1987. xi, 488 p. $14.95.

BERNSTEIN, RICHARD J., ed.: *Habermas and Modernity.* Cambridge: MIT Press, 1985. viii, 243 p. $8.95.

BURGER, HEINZ, ed.: *Zeit Natur und Mensch.* Berlin: Berlin Verlag, 1986. 372 p. 38 DM.

CAMPBELL, RICHMOND, and LANNING SOWDEN, eds.: *Paradoxes of Rationality and Cooperation: Prisoner's Dilemma and Newcomb's Problem.* Vancouver: British Columbia UP, 1985. 366 p. $19.95.

COHEN, ROBERT S., and MARX W. WARTOFSKY, eds.: *A Portrait of Twenty-five Years: Boston Studies in the Philosophy of Science 1960–1985.* Boston: D. Reidel, 1985. viii, 326 p. $14.95.

DAVIS, MICHAEL, and FREDERICK A. ELLISTON, eds.: *Ethics and the Legal Profession.* New York: Prometheus, 1986. v, 495 p. $16.95.

HABERMAS, JÜRGEN, ed.: *Observations on "The Spiritual Situation of the Age."* ANDREW BUCHWALTER, trans. Cambridge, Mass.: MIT Press, 1986. xxxvii, 381 p. $10.95.

HAHN, LEWIS EDWIN, and PAUL ARTHUR SCHILPP, eds.: *The Philosophy of W. V. Quine.* La Salle, Ill.: Open Court, 1986. xvi, 705 p. $19.95.

HARRISON, STANLEY M., and RICHARD C. TAYLOR, eds.: *The Life of Religion: A Marquette University Symposium on the Nature of Religious Belief.* Lanham, Md.: University Press of America, 1986. xxxi, 92 p. $8.75.

INWOOD, MICHAEL, ed.: *Hegel.* New York: Oxford, 1985. 217 p. $7.95.

KOURANY, JANET A., ed.: *Scientific Knowledge.* Belmont, Calif.: Wadsworth, 1987. x, 399 p. $23.

LOCKWOOD, MICHAEL, ed.: *Moral Dilemmas in Modern Medicine.* New York: Oxford, 1986. viii, 259 p. $9.95.

MCCORMICK, PETER J., ed.: *The Reasons of Art.* Ottawa: University Press, 1986. xviii, 496 p. $34.95.

MEYER, SAMUEL, ed.: *Dewey and Russell: An Exchange.* New York: Philosophical Library, 1985. 112 p. $9.95.

NEUMAIER, OTTO, ed.: *Mind, Language and Society.* Vienna: VWGÖ, 1984. 229 p.

NYÍRI, J. C., ed.: *From Bolzano to Wittgenstein: The Tradition of Austrian Philosophy.* Boston: D. Reidel, 1986. 200 p. $38.

PAUL, ELLEN FRANKEL, FRED D. MILLER, JR., and JEFFREY PAUL, eds.: *Ethics and Economics.* New York: Basil Blackwell, 1985. 191 p. $11.95.

PAUL, ELLEN FRANKEL, FRED D. MILLER, JR., and JEFFREY PAUL, eds.: *Human Rights.* New York: Basil Blackwell, 1986. 175 p. $12.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

PHILLIPS, D. C., and JONAS F. SOLTIS, eds: *Perspectives on Learning*. New York: Columbia, 1985. vii, 98 p. $8.95.

RESCHER, NICHOLAS, ed.: *Current Issues in Teleology*. Lanham, Md.: University Press of America, 1986. vii, 135 p. $9.25.

RESCHER, NICHOLAS, ed.: *Reason and Rationality in Natural Science*. Lanham, Md.: University Press of America, 1985. 219 p. $12.25.

STEWART, ROBERT M., ed.: *Readings in Social and Political Philosophy*. New York: Oxford, 1986. viii, 420 p. $16.95.

NEW BOOKS IN PAPERBACK

ASHMORE, ROBERT B.: *Building a Moral System*. Englewood Cliffs, N.J.: Prentice-Hall, 1986. xi, 175 p. $16.95.

BERWICK, ROBERT C., and AMY S. WEINBERG: *The Grammatical Basis of Linguistic Performance*. Cambridge, Mass.: MIT Press, 1986. xviii, 325 p. $9.95.

BOHM, PETER: *Theodor Lessings Versuch einer erkenntnistheoretischen Grundlegung von Welt*. Atlantic Highlands, N.J.: Humanities Press, 1986. 127 p. $19.95.

BRAYBROOKE, DAVID: *Philosophy of Social Science*. Englewood Cliffs, N.J.: Prentice-Hall, 1987. vii, 136 p. $14.95.

BROSE, KARL: *Sprachspiel und Kindersprache: Studien zu Wittgensteins "Philosophischen Untersuchungen"*. Frankfurt/Main: Campus Forschung, 1985. 336 p. 58 DM.

BROWN, ALAN: *Modern Political Philosophy*. New York: Viking/Penguin, 1986. 215 p. $6.95.

BROWN, ROBERT: *The Nature of Social Laws: Machiavelli to Mill*. New York: Cambridge, 1986. ix, 270 p. $12.95.

CAMPBELL, KEITH: *A Stoic Philosophy of Life*. Lanham, Md.: University Press of America, 1986. xi, 201 p. $12.25.

CLASSEN, CARL JOACHIM: *Ansätze*. Atlantic Highlands, N.J.: Humanities, 1986. 288 p. $39.95.

COX, J. GRAY: *The Ways of Peace: A Philosophy of Peace as Action*. Mahwah, N.J.: Paulist Press, 1986. v, 211 p. $11.95.

CRAMER, KONRAD: *Nicht-reine synthetische Urteile a prior: Ein Problem der Transzendentalphilosophie Immanuel Kants*. Heidelberg: Carl Winter, 1985. 385 p. 94 DM.

DAVIGNON, RENE: *Le Mal chez Gabriel Marcel: Comment affronter la souffrance et la mort?* Montreal: Editions Bellarmin, 1985. 174 p. $12.

DENNETT, D. C.: *Content and Consciousness*. New York: Methuen, 1986. xvi, 198 p. $12.95.

DILMAN, ILHAM: *Freud and Human Nature*. New York: Basil Blackwell, 1986. vi, 207 p. $9.95.

FRITSCHE, JOHANNES: *Methode und Beweisziel*. Meisenheim am Glan: Anton Hain, 1986. 268 p. 54 DM.

HALLER, RUDOLF: *Fragen zu Wittgenstein*. Atlantic Highlands, N.J.: Humanities, 1986. 254 p. $35.

HARTKOPF, WERNER: *Studien zu Schellings Dialektik*. Meisenheim am Glan: Anton Hain, 1986. 330 p. 78 DM.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

IGNATIEFF, MICHAEL: *The Needs of Strangers.* New York: Viking/Penguin, 1985. 156 p. $6.95.

JUHL, P.D.: *Interpretation: An Essay in the Philosophy of Literary Criticism.* Princeton, N.J.: University Press, 1986. x, 332 p. $12.50.

KÖRNER, STEPHAN: *Metaphysics: Its Structure and Function.* New York: Cambridge, 1986. xi, 238 p. $14.95.

KUHSE, HELGA, and PETER SINGER: *Should the Baby Live? The Problem of Handicapped Infants.* New York: Oxford, 1985. vii, 228 p. $17.95.

LAUDAN, LARRY: *Science and Values: The Aims of Science and Their Role in Scientific Debate.* Berkeley: California UP, 1986. xiv, 149 p. $8.95.

MAGEE, BRYAN: *Modern British Philosophy.* New York: Oxford, 1986. 287 p. $9.95.

MICHAUD, YVES: *Locke.* Paris: Bordas, 1986. 189 p.

NYIRI, J. C.: *Gefühl and Gefüge.* Atlantic Highlands, N.J.: Humanities, 1986. 207 p. $29.95.

REEDER, HARRY P.: *The Theory and Practice of Husserl's Phenomenology.* Lanham, Md.: University Press of America, 1986. xiii, 186 p. $11.75.

REITH, HERMAN R.: *René Descartes: The Story of a Soul.* Lanham, Md.: University Press of America, 1986. xiii, 197 p. $12.50.

RESCHER, NICHOLAS: *Ongoing Journey.* Lanham, Md.: University Press of America, 1986. ix, 231 p. $13.75.

ROHATYN, DENNIS: *The Reluctant Naturalist.* Lanham, Md.: University Press of America, 1987. viii, 141 p. $9.75.

ROSEN, STANLEY: *Plato's* Sophist: *The Drama of Original and Image.* New Haven: Yale, 1986. x, 341 p. $14.95.

SALMON, NATHAN: *Frege's Puzzle.* Cambridge, Mass.: MIT Press, 1986. xi, 194 p. $12.50.

SMYTH, JOHN VIGNAUX: *A Question of Eros: Irony in Sterne, Kierkegaard, and Barthes.* Tallahassee: Florida State UP, 1986. xv, 415 p. $29.95.

STEGMÜLLER, WOLFGANG: *Probleme und Resultate der Wissenschaftstheorie und analytischen Philosophie. Band II: Theorie und Erfahrung.* Berlin: Springer-Verlag, 1986. Teil F, x, 136 p., 40 DM. Teil G, 137–310 pp., 45 DM. Teil H, 311–460 p., 40 DM.

TALBUTT, PALMER, JR.: *Reanimation in Philosophy.* Lanham, Md.: University Press of America, 1986. xii, 207 p. $11.50.

THEAU, JEAN: *Le Crépuscule de l'homme.* Montreal: Bellarmin, 1986. 161 p. $15.

VERNANT, DENIS: *Introduction à la philosophie de la logique.* Brussels: Pierre Mardaga, 1986. 255 p. 1.250 FB.

YOUNG, THEODORE A.: *Completing Berkeley's Project: Classical vs. Modern Philosophy.* Lanham, Md.: University Press of America, 1985. xi, 213 p. $11.75.

WEBER, RENÉE: *Dialogues with Scientists and Sages: The Search for Unity.* Boston: Routledge & Kegan Paul, 1986. xiii, 256 p. $14.95.

WILSON, JOHN: *What Philosophy Can Do.* Totowa, N.J.: Barnes & Noble, 1986. viii, 159 p. Cloth $23.50, paper $9.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NEW ANTHOLOGIES

ACHINSTEIN, PETER, and OWEN HANNAWAY, eds.: *Observation, Experiment, and Hypothesis in Modern Physical Science*. Cambridge, Mass.: MIT Press, 1985. x, 379 p. $35.

ADAMS, JAMES LUTHER, WILHELM PAUCK, and ROGER LINCOLN SHINN, eds.: *The Thought of Paul Tillich*. San Francisco: Harper & Row, 1985. xi, 404 p. $24.95.

ANNAS, JULIA, ed.: *Oxford Studies in Ancient Philosophy*. New York: Oxford. Vol. I, 1984: ix, 257 p.; cloth $39.95, paper $19.95. Vol. II, 1984: 303 p.; cloth $39.95, paper $19.95. Vol. III, 1985: ix, 338 p.; cloth $24.95, paper $14.95.

AUDI, ROBERT, and WILLIAM J. WAINWRIGHT, eds.: *Rationality, Religious Belief, and Moral Commitment*. Ithaca, N.Y.: Cornell, 1986. 341 p. Cloth $42.50, paper $12.95.

BEVER, THOMAS G., JOHN M. CARROLL, and LANCE A. MILLER, eds: *Talking Minds: The Study of Language in the Cognitive Sciences*. Cambridge, Mass.: MIT Press, 1985. 283 p. $19.95.

BOGDAN, RADU J., ed.: *Belief: Form, Content and Function*. New York: Oxford, 1986. vi, 184 p. Cloth $42, paper $15.95.

BOGDAN, RADU J., ed.: *Roderick M. Chisholm*. Boston: D. Reidel, 1986. ix, 237 p. $59.50.

BRODSKY, GARRY, JOHN TROYER, and DAVID VANCE: *Contemporary Readings in Social and Political Ethics*. Buffalo, N.Y.: Promethus, 1985. x, 436 p. $15.95.

BULYGIN, EUGENIO, JEAN-LOUIS GARDIES, and ILKKA NIINILUOTO, eds. *Man, Law and Modern Forms of Life*. MICHAEL D. BAYLES, trans. Boston: D. Reidel, 1985. 321 p. $39.50.

BUTTERFIELD, JEREMY, ed.: *Language, Mind, and Logic*. New York: Cambridge, 1986. xi, 232 p. $34.50.

CAMPBELL, TOM, DAVID GOLDBERG, SHEILA McLEAN, and TOM MULLEN, eds.: *Human Rights: From Rhetoric to Reality*. New York: Basil Blackwell, 1986. 262 p. Cloth $45, paper $19.95.

CARRITHERS, MICHAEL, STEVEN COLLINS, and STEVEN LUKES, eds.: *The Category of the Person*. New York: Cambridge, 1985. viii, 309 p. Cloth $39.50, paper $13.95.

CASEY, EDWARD S., and DONALD V. MORANO: *The Life of the Transcendental Ego: Essays in Honor of William Earle*. Albany: SUNY Press, 1986. xii, 217 p. Cloth $36.50, paper $12.95.

CHURCHLAND, PAUL M., and CLIFFORD A. HOOKER, eds.: *Images of Science*. Chicago: University Press, 1985. viii, 309 p. cloth $45, paper $18.95.

COHEN, AVNER, and STEVEN LEE, eds.: *Nuclear Weapons and the Future of Humanity: The Fundamental Questions*. Totowa, N.J.: Rowman & Allanheld, 1986. xii, 496 p. Cloth $39, paper $16.95.

COHEN, RICHARD A., ed.: *Face to Face with Levinas*. Albany: SUNY Press, 1986. viii, 264 p. Cloth $39.50, paper $12.95.

COHEN, ROBERT S., RICHARD M. MARTIN, and MEROLD WESTPHAL, eds.: *Studies in the Philosophy of J. N. Findlay*. Albany: SUNY Press, 1985. xv, 478 p. Cloth $49.50, paper $23.50.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIV, NO. 4, APRIL 1987

FUNCTIONS

IN describing the function of some biological character, we describe some presently existing item by reference to some future event or state of affairs. So the function of teeth at time t is to pulp food at time t', where $t' > t$. This seems to present an exact parallel to the case of the function of humanmade artifacts; for instance, the function of the nutcracker at time t is to break open nuts at time t', where $t' > t$.

In the case of biological functions and in other cases of functions where human intentions are not obviously causally active, this kind of description of function has been difficult to assimilate into our scientific view of the world. There are several reasons, but we shall here concentrate on one which arises directly from the fact that, in describing a present structure in terms of its function, we mention a future outcome of some sort. The future outcome may be, in many cases, nonexistent. A structure may never be called upon to perform that function. The function of a bee's sting, for instance, is relatively clear; yet most bees never use their stings. Likewise teeth may never pulp food, just as nutcrackers may never crack nuts.

Thus, when we describe the function of something in the present, we make reference to a future event or effect which, in some cases, will never occur. Hence, prima facie, we cannot really be describing any genuine, current property of the character.

I. THE PROBLEM

Even when a character does perform its supposed function, the future events that result from it cannot play any significant "scientific" role in *explaining* the nature and existence of the character. The character has come into existence, and has the properties that it does have, as a result of prior causes. It would still have existed, with just the current properties it does have, even if it had not been followed by the events that constitute the exercise of its alleged

0022-362X/87/8404/0181$01.60 © 1987 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

function. Hence its existence and properties do not depend on the exercising of its function. So it is hard to see what explanatory role its functions could have. Crudely put: backwards causation can be ruled out—structures always have prior causes—hence reference to future events is explanatorily redundant. Hence functions are explanatorily redundant.

Of course, there is nothing inappropriate about describing a character *and mentioning* its future effects. But describing a character as having a function is not just mentioning that it has certain effects. Not every effect counts as part of its function. And some functions are present even when there are no relevant effects to be mentioned —as with some bees and their stings. Future events are not unmentionable; but they are explanatorily redundant in characterizing the existence and current properties of a character. Hence, what role can functions have in a purely scientific description of the world: how can they be "placed" within the framework of current science?

There are three main theories that attempt to construe functions in a way that allows them to fit smoothly into the scientific, causal order. We believe that each is nearly right or partly right. Yet they are all unsatisfactory in one crucial respect: they do not restore to functions any significant explanatory power. In particular, they deny to functions any causal efficacy. So, for instance, they will not permit us to explain the evolution of a character by saying that it evolved *because* it serves a specific function.

We will offer an account of biological function and of functions generally which, although it shares much with the most promising extant theories, is crucially different from them in that it bestows greater explanatory power upon functions.[1] But first we will briefly consider the three theories.

II. ELIMINATIVISM

There are three responses that arise naturally in the face of the tension between functions and the scientific standpoint. The first is eliminativist. It is assumed that functions, if there were any, would have to be important, *currently existing,* causally active, and explanatory properties of a character or structure. It is also assumed that functions, if there were any, would essentially involve reference to

[1] Our theory will be a cousin of theories which are sometimes called "goal theories" and which have been advocated, for instance, by Christopher Boorse, "Wright on Functions," *Philosophical Review,* LXXXV, 1 (January 1976): 70–86, and "Health as a Theoretical Concept," *Philosophy of Science,* XLIV, 4 (December 1977): 542–573; and by William Wimsatt, "Teleology and the Logical Structure of Function Statements," *Studies in the History and Philosophy of Science,* III, 1 (May 1972): 1–80. Some key differences between their theories and ours will be noted later.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

future, possibly nonexistent, events. Yet something involving essential reference to future, possibly nonexistent events could not possibly characterize currently existing, explanatory properties. It is thus concluded that there really are no functions in nature.

To add functions to the scientific biological picture, on this view, is parallel to adding final causes to physics. Final causes have no place in the scientific account of the physical universe, and, if the psychological pressures are resisted, we find we can do without them and final causes just fade away. To the eliminativist, the same will be true of functions; as the biological sciences develop, any need for function talk will vanish, and the psychological naturalness of such talk will fade away with time and practice.

A variant on this eliminativist view adds an account of why the attributions of function seem to serve a useful purpose in everyday and scientific discourse. The eliminativist does not believe in functions as genuine, currently existing properties of a character. But the eliminativist does believe in future effects of a character. And nothing stops us from mentioning whichever future effects we take an interest in. Consequently, an eliminativist can interpret talk of "functions" as being merely the specification of effects one happens to be interested in. Which effects are deemed to relate to "functions" of a character, will depend not on the nature of the character itself, but on our interests. The function of kidneys is different for the anatomist from what it is for the chef. Insofar as function talk makes sense, it does *not* describe the current nature of a character (there are no functions in nature); rather, it relates a current character to a future outcome, in an interest-dependent, extrinsic manner.[2]

The best answer to an eliminativist theory is to come up with an adequate analysis of functions still within the scientific view. This is what we attempt to do later in this paper.

But a motive for seeking such a noneliminativist account can be cited; and this motive will also provide a less than conclusive, but nevertheless weighty, argument against eliminativism. In the biological sciences, functions are attributed to characters or structures, and these attributions are intended to play an explanatory role which cannot be squared with the eliminativist's account of function talk. For instance it is assumed that biological structures *would* have had the functions they do have even if we had not been here to take an interest in them at all. And some of the effects of structures that we

[2] A theory of this sort can be found in Robert Cummins, "Functional Analysis," this JOURNAL, LXXII, 20 (Nov. 20, 1975): 741–765. We have also been influenced by a paper by Elizabeth W. Prior, "What Is Wrong with Etiological Accounts of Biological Function?," *Pacific Philosophical Quarterly*, LXVI, 3/4 (July/October 1985): 310–328.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

take an interest in have nothing to do with their function. And some functions are of no interest to us at all. Furthermore, biology standardly treats function as a central, explanatory concept. None of this rests easily with an eliminativist theory. A powerful motive for resisting eliminativism, then, is that an adequate analysis of functions, if we can find one, will enable us to take much biological science at face value: we will be relieved of the necessity of undertaking a radical reformation of the biological sciences. The eliminativists' vision, of functions "fading away," is as yet just a pipe dream; and their explanation of the apparent usefulness of function talk fails to explain away more than a fragment of the uses functions serve in biological science.

It is not the biological sciences alone which could be cited here. Psychology could be canvassed too. And even physics has facets that raise problems for an eliminativist view. Suppose someone were to suggest that the function of water is to refract light and the function of mists to create rainbows. Presumably this is plainly false. Yet it describes something in terms of future effects in which we take an interest . . . which is exactly what the eliminativist takes to be the business of function talk. So eliminativists have no good explanation of why the physicist takes such attributions of function to be plainly false. They cannot explain the manifest difference between "The function of mists is to make rainbows," and "The function of teeth is to pulp food."[3]

III. REPRESENTATIONAL THEORIES

There is a response to the tension between the scientific view and functions which rejects any role for future events in the characterizing of functions. The future effects of a character do not themselves play an explanatory role in characterizing that character. Yet sometimes there exists, prior to the character, a *plan,* a representation of that character *and* of its future effects. Such a representation of future effects may exist, whether or not those effects ever come to pass. And this representation exists prior to the character and so contributes to the causal processes that bring that character into being—by the usual, forward-looking, causal processes that rest comfortably within our over-all scientific image of the world. On this view, we can account for a function not by direct reference to any

[3] This problem is treated at some length by Ernest Nagel, *The Structure of Science* (New York: Harcourt; London: Routledge, 1961), in his section on teleology, pp. 401–428. Nagel manages to blunt the force of such objections, but only by augmenting his initial theory, which differs from eliminativism only superficially, thereby generating a theory that comes close to the theories of Boorse and Wimsatt mentioned above.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

future event, but rather by reference to a past representation of a future event. Theories of this sort have frequently been called "goal theories"; but they are best construed as a subcategory within the class of goal theories. They are goal theories in which the identification of a goal depends on the content of prior representations.[4]

This kind of account of functions fits best with attribution of functions to artifacts. The idea of breaking open nuts seems to have played a causal role in the production of the nutcracker. It does not fit so neatly into the biological sciences. Of course, it used to provide a persuasive argument (the teleological argument) for the existence of a Creator: there *are* functions in nature; functions require prior representations; yet the creatures themselves (even when creatures are involved) have no such foresight or were not around at the right time; hence the prior representations must have been lodged in some awfully impressive being . . . etc. Nowadays, however, in the clear and noncontroversial cases of functions in nature, it is taken that they can be accounted for from the standpoint of a theory of evolution by way of *natural* selection—and in such a theory there can be no room for any analysis of biological functions which rests on prior representations. Even if God foresaw the functions of biological structures, that is a matter outside biology; functions, however, are a biological and not a theological matter.

It is worth noting that, even though the representational theory seems to rest comfortably with attributions of functions to artifacts, nevertheless, some artifacts prove more problematic than might first appear. Many artifacts evolve by a process very like natural selection. Variations often occur by chance and result in improved performance. The artisan may not understand fully the reasons why one tool performs better than others. Yet, because it performs well, it may be copied, as exactly as possible. The reproduction of such tools may occur for generations. The features of the tool which make it successful and which lead it to be selected for reproduction are features that have specific functions. But they were not created with those functions in mind. They may have been produced with an over-all function in mind (say, hitting nails); but the toolmaker may not have in mind any functions for the components and features of the tool, which contribute to the over-all function. For instance, the toolmakers may copy a shape that has the function of giving balance to the tool—but they need not foresee, or plan, or represent any

[4] Andrew Woodfield, *Teleology* (New York: Cambridge, 1976), argues for a view that takes the primary cases of functions to rest on a prior plan, and all other cases of (unplanned) "functions" to be mere metaphorical extensions of the primary cases.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

such function. They know only that tools like this work well at banging nails, or sawing wood, or whatever the over-all function might be. Consequently, even with artifacts, structures can serve specific functions even though there exists no prior representation of that function.[5]

There is a further reason for uneasiness about the representational theory. The theory analyzes the *apparent* forward directedness of functions by an indirect, two-step, route. The forward directedness of functions is analyzed as comprising a *backward* step to a representation, which in turn has a *forward* directedness toward a possibly nonexistent future state.

Thus the seeming forward directedness of functions is reduced to another sort of forward directedness: that of representations—plans, beliefs, intentions, and so on. And this is worrying. The worry is not just that these are "mentalistic," and just as problematic as functions—and just as hard to assimilate into the scientific picture of the world. Rather, an even greater worry for many will be that of vicious circularity. Many find it plausible that the notion of representation will turn out to be analyzable in terms that at least include functional terms. And functional terms presuppose functions. Hence the future directedness of representations may turn out to presuppose the future directedness of functions. This threatens to do more than just restrict the scope of representational theories; it undermines such theories even in their home territory, as applied to artifacts.

IV. ETIOLOGICAL THEORIES

The third response to the tension between the scientific view of the world and the concept of function again involves rejecting any role for future events in the characterization of functions. Yet etiological theories also eschew reference to prior representations of future effects, as well as reference to the future effects themselves.[6]

Representational theories and etiological theories have an important feature in common. Both shun any genuine, direct reference to

[5] There are several intriguing points made about artifacts, their reproduction, selection, survival, and so forth, by Ruth Millikan, *Language, Thought and Other Biological Categories: New Foundations for Realism* (Cambridge, Mass.: MIT Press, 1984). Millikan advances a sophisticated version of the etiological theory, which we discuss below.

[6] The etiological theory is widely held, but a very good exposition and defense is given by Larry Wright, "Functions," *Philosophical Review*, LXXXII, 2 (April 1973): 139–168; and *Teleological Explanations* (Los Angeles: California UP, 1976). We have also been greatly influenced by the defense of etiological theories advanced by Karen Neander, *Abnormal Psychobiology*, Ph.D. thesis, La Trobe University, 1983.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

future effects, and refer instead only to past causes. Both construe the attribution of function as supplying information about the genesis of the character, that is, about how the character came into existence.

The difference between the two theories recalls the distinction Charles Darwin drew between artificial selection and natural selection. When animal breeders select, they represent to themselves the characters they wish to develop. Natural selection has closely analogous results, but it operates in the absence of (or at least without any need for) any conscious or unconscious representations of future effects.

The etiological theory of functions explains biological functions by reference to the process of natural selection. Roughly: a character has a certain function when it has evolved, by natural selection, *because* it has had the effects that constitute the exercise of that function.

Clearly, there is room here for an overarching, disjunctive theory, which unites the representational with the etiological. Such an overlapping theory would say: a character has a certain function when it has been selected because that character has had the relevant effects. In the case of artifacts, the selection involves conscious representations (mostly); in the case of (Darwinian) sexual selection, representations may enter the picture, but reproduction, heredity, and evolution also play a part; and in the case of natural selection, representations drop out altogether.

But, on the etiological theory, a character has a biological function only if that character has been selected for by the process of natural selection because it has had the effects that constitute the exercise of that function. This is the only kind of selection compatible with the dictates of modern biological science.

We take this etiological theory of biological function as the main alternative to the account of biological function we shall proffer. The big plus for the etiological theory is that it makes biological functions genuinely explanatory, and explanatory in a way most comfortable with the modern biological sciences. But we shall argue that this explanatory power is still not quite right: it offers explanations that are too backward-looking. The theory we offer will be more forward-looking in its explanatory nature.

But, before we turn to this matter, we should note another worry with the etiological theory, a worry that extends even to the overarching disjunctive theory of which it is a part. This worry is that there is too great a dependence of the intrinsic nature of functions on contingent matters—matters which, had they been (or even if

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

they are) otherwise, would rob the theory of any viability as an account of functions.

The etiological theory of biological functions has such functions characterized in terms of evolution by natural selection. Most theorists take evolutionary theory to be true, but contingently so. What if the theory of evolution by natural selection were to be (or had been) false? Clearly then, on the etiological theory of biological functions, as we have specified it, there would be no biological functions. Whether or not there are biological functions at all, on the over-arching, disjunctive theory, will depend on what replaces the theory of evolution by natural selection. Suppose it is creationism. Then the representational theory would apply; for we have the representations in the mind of the creator that would have the appropriate causal role in the development of biological structures, so the representational theory would become a general theory of functions.

We noted earlier that the representational theory had problems with the functions of some artifacts: artifacts that seemed to evolve over time by a process similar to natural selection. If this analogous process is also not available, along with natural selection proper, then our representational theory will not bestow functions upon such artifacts. But perhaps this is small change.

For creationism, there would of course be an enormous epistemological problem of discovering what the functions of biological structures were; for this would depend on discovering what the creator had in mind. So we would be stuck with great difficulty in discovering whether the function of the heart is to produce the sound of a heart beat, in line with the creator's idea of a beating rhythm in nature, with the circulation of the blood as a nonfunctional effect; or whether the reverse is true. It would be much like an anthropologist discovering the nature of ancient artifacts without any presuppositions about the intentions of the earlier cultures.

But suppose creationism is not the alternative. Consider the possible world identical to this one in all matters of laws and particular matters of fact, except that it came into existence by chance (or without cause) five minutes ago. Now, on even the over-arching theory, there are no functions; for there are no biological functions on the etiological theory, and no causally active representations as required by the representational theory, and hence no functions at all.

We have the intuition that the concept of biological function, and views about what functions biological characters have, are not thus contingent upon the acceptance of the theory of evolution by natural selection and on discovering what led to the evolutionary develop-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ment of particular characters. In parallel, we are also inclined to think that the representational account of the function of artifacts gives too much importance to the representations or ideas of the original planner, even in cases when there is one. As we indicated earlier, we believe a satisfactory account of functions in general, and of biological functions in particular, must be more forward-looking. We now turn to this.

V. FITNESS, FUNCTION, AND LOOKING FORWARD

It emerges from our discussions that the tension between functions and modern, causal science has generated, fundamentally, two stances on the nature of functions.

The first is the eliminativist stance. This has the merit of giving full weight to the forward-looking character of functions, by specifying them in terms of future and perhaps nonexistent effects; and also to the explanatory importance of functions. It is mistaken only in its despair of reconciling these two strands.

The second stance is backward-looking. This embraces theories which look back to prior representations and those which look back to a prior history of natural selection and those which look back to a history of either one sort or the other.

We will argue for a forward-looking theory. Functions can be characterized by reference to possibly nonexistent future events. Furthermore, they *should* be characterized that way, because only then will they play the explanatory role they need to play, for instance, in biology. The way to construe functions in a forward-looking manner, we suggest, is (roughly) to construe them in the manner of dispositions. The shift we recommend, in our conception of functions, has a precedent: the analysis of the evolutionary concept of fitness.

One wrongheaded, but at times common, objection to the Darwinian theory of evolution is that its central principle—roughly, "the survival of the fittest"—is an empty tautology which cannot possibly bear the explanatory weight Darwin demands of it.[7] This objection

[7] For those who have suggested this view, with greater or less refinement and sophistication, see J. J. C. Smart, *Philosophy and Scientific Realism* (New York: Random House, 1963), p. 59; H. G. Cannon, *The Evolution of Living Things* (Manchester: University Press, 1958); C. H. Waddington, *The Strategy of the Genes* (London: Allen & Unwin, 1957), pp. 64/5; A. O. Barker, "An Approach to the Theory of Natural Selection," *Philosophy*, XLIV, 170 (October 1969): 271–290; Robert Brandon and John Beatty, "The Propensity Interpretation of 'Fitness': No Interpretation Is No Substitute," *Philosophy of Science*, LI, 2 (June 1984): 342–347. For those who have replied to this view, see in particular Edward Manier, " 'Fitness' and Some Explanatory Patterns in Biology," *Synthese*, XX, 2 (August 1969): 206–218; and Michael Ruse, "Natural Selection in the *Origin of Species*," *Studies in the History and Philosophy of Science*, I (February 1971): 311–351.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

assumes that fitness can be judged only retrospectively: that it is only after we have seen which creatures survived that we can judge which were the fittest; moreover, it assumes that the fact that certain creatures have survived, whereas others did not, is what constitutes their being the fittest.

The etiological theory of biological functions rests on the same sort of misconception as that which underlies the vacuity objection to Darwin. On this theory, we can judge only retrospectively that a character has a certain function, when its having had the relevant effect has contributed to survival. Indeed, on the etiological theory, that an effect is part of the function of a character is *constituted* by the fact that having this effect has contributed to the survival of the character and of the organisms that bear it.

Consequently, the notion of function is emptied of much explanatory potential. It is no longer possible to explain why a character has persisted by saying that the character has persisted because it serves a given function. To attempt to use function in that explanatory role, would be *really* to fall into the sort of circularity often alleged (falsely) against the explanatory use of fitness in Darwinism.

This comparison with fitness serves another purpose. It has displayed why functions would lack explanatory power on the etiological theory, but it also shows how to analyze functions so as not to lose this explanatory power. Fitness is not defined retrospectively, in terms of actual survival. It is, roughly, a dispositional property of an individual (or species) in an environment, which bestows on that individual (or species) a certain survival potential or reproductive advantage. This is a subjunctive property: it specifies what will happen or what is likely to happen in the right circumstances, just as fragility is specified in terms of breaking or being likely to break in the right circumstances. And such a subjunctive property supervenes on the morphological characters of the individual (or species).[8] Hence there is no circularity involved in casting fitness in an explanatory role in the Darwinian theory of evolution. In the right circumstances fitness explains actual survival or reproductive advantage, just as in the right circumstances fragility explains actual breaking. In each case the explanation works by indicating that the individual has certain causally active properties that in such circumstances will bring about the phenomena to be explained.

What holds here of fitness holds, too, of biological functions. The

[8] For more on dispositions and their supervenience on categorical bases, see Pargetter and Prior, "The Dispositional and the Categorical," *Pacific Philosophical Quarterly*, LXIII, 4 (October 1982): 366–370; and Prior, Pargetter, and Frank Jackson, "Three Theses about Dispositions," *American Philosophical Quarterly*, XIX, 3 (July 1982): 251–257.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

etiological theory is mistaken in defining functions purely retrospectively, in terms of actual survival. Hence there need be no circularity in appealing to the functions a character serves, in explaining the survival of the character. Fitness is forward-looking. Functions should be forward-looking in the same way and, hence, are explanatory in the same way.

VI. THE PROPENSITY THEORY

Here is one way to derive a "forward-looking" theory of functions.

Let us begin with the etiological theory. Consider a case in which some character has a specific effect and has been developed and sustained by natural selection because it had that effect. In such a case the etiological theory deems that it is (now) a function of the character to produce that effect.

Look more closely, then, at the past process that has "conferred" a function, according to the etiological theory. The character in question must have had the relevant effect, on a sufficient number of occasions—and in most cases, this will have been not on randomly chosen occasions, but on *appropriate* occasions, in a sense needing further clarification. (For instance, sweating will have had the effect of cooling the animal—and it will have had this effect on occasions when the animal was hot, not when it was cold.)

The history that confers a function, according to the etiological theory, will thus display a certain pattern. The effect that will eventually be deemed a function must have been occurring in appropriate contexts; that is to say, it must have been occurring in contexts in which it contributes to survival, at least in a statistically significant proportion of cases.

Further, this contribution to survival will not, in realistic cases, have been due to sheer accident. One can imagine individual incidents in which a character contributes to survival by sheer chance. Laws of probability dictate that a long run of such sheer accidents is conceivable; but it will be very unlikely—except for characters with relatively short histories. The only cases in which such long runs are likely to occur are cases in which the character confers a standing propensity upon the creature, a propensity that increases its chances of survival.

If we imagine (or find in the vast biological record) a case in which a character is sustained by a chance sequence of accidents, rather than by a standing propensity, then it would not be appropriate to describe that character as having a function. This can happen when a character is linked with another character that does bestow a propensity and where variations in the character just have not occurred to allow selection against the inoperative character. It can also happen by sheer chance, a long-run sequence of sheer flukes. Such a

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sequence is very improbable; but biology offers a stunningly large sample. It is very probable that many improbable events will have occurred in a sample that large.

Consequently, what confers the status of a function is not the sheer fact of survival-due-to-a-character, but rather, survival due to the propensities the character bestows upon the creature.

The etiological theory describes a character *now* as serving a function, when it *did* confer propensities that improved the chances of survival. We suggest that it is appropriate, in such a case, to say that the character *has been serving that function all along*. Even before it had contributed (in an appropriate way) to survival, it had conferred a survival-enhancing propensity on the creature. And to confer such a propensity, we suggest, is what constitutes a function. Something has a (biological) function just when it confers a survival-enhancing propensity on a creature that possesses it.

Four features of this propensity theory of biological functions should be made explicit.

First, like the corresponding account of fitness, this account of functions must be relativized to an environment. A creature may have a high degree of fitness in a specific climate—but a low degree of fitness in another climate. Likewise, a character may confer propensities which are survival-enhancing in the creature's usual habitat, but which would be lethal elsewhere. When we speak of the function of a character, therefore, we mean that the character generates propensities that are survival-enhancing in the creature's natural habitat. There may be room for disagreement about what counts as a creature's "natural habitat"; but this sort of variable parameter is a common feature of many useful scientific concepts.

Ambiguities will arise especially when there is a sudden change in the environment. At first, we will refer the creature's "natural habitat" back to the previous environment. But eventually we will transfer the term to the current environment. The threshold at which we make such a transference will be vague. The notion of natural habitat will also be ambivalent as applied to domestic animals.

In its most obvious use, the term 'habitat' applies to the physical surroundings of a whole organism. But we can also extend its usage, and apply the term 'habitat' to the surroundings of an organ within an organism. Or to the surroundings of a cell within an organ. In each case, the natural habitat of the item in question will be a functioning, healthy, interconnected system of organs or parts of the type usual for the species in question. When some of the organs malfunction, then other organs, which go on performing their natural functions, may no longer be contributing to survival. We still say they are

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

performing their natural function, even though this does not enhance the chance of survival. Why? Because it would enhance survival if the other organs were performing as they do in healthy individuals.

Consequently, functions can be ascribed to components of an organism, in a descending hierarchy of complexity. We can select a subsystem of the organism, and we can ascribe a function to it when it enhances the chances of survival in the creature's natural habitat. Within this subsystem, there may be a subsubsystem. And this may be said to serve a function if it contributes to the functioning of the system that contains it—provided all other systems are functioning "normally" (that is, provided it is lodged in its own "natural habitat"). And so on.

Similar hierarchies may also occur in the opposite direction: a microscopic organism has a function in a pond, which has a function in a forest, which has a function in the biosphere, and so on.

Secondly, on the propensity theory, functions are truly dispositional in nature. They are specified subjunctively: they *would* give a survival-enhancing propensity to a creature in an appropriate manner, in the creature's natural habitat.[9] This is true even if the creature does not survive or is never in its natural habitat. Likewise, fragility gives a propensity to break to an object in an appropriate manner, in the right circumstances—and of course some fragile objects never break. And fitness gives a propensity to an individual or species to survive, in an appropriate manner, in a specified environment and in a struggle for existence, even if there is no struggle for existence or if the individual or species fails to survive.

Of course, when functions do lead to survival—just as when dispositions are manifested—the cause will be the morphological structural form of the creature and the relationship between this form and the environment. Functions supervene on this in the same way that dispositions supervene on their categorical bases. But the functions will be explanatory of survival, just as dispositions are explanatory of their manifestations; for they will explain survival by pointing to the existence of a character or structure in virtue of which the creature has a propensity to survive.

Thirdly, in the long run, it will be necessary to spell out the notion

[9] It is this central role we give to propensities which distinguishes our theory from others, like those of Boorse and Wimsatt (mentioned above), which fall back on the notion of statistically normal activities within a class of organisms. On their theories, a character has a function for a creature when it *does* help others "of its kind" to survive, in a sufficiently high proportion of cases. On our view, frequencies and statistically normal outcomes will be important evidence for the requisite propensities. But there are many well-known and important ways in which frequencies may fail to match propensities.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of a "survival-enhancing propensity" in formal terms, employing the rigors of the probability calculus. Clearly, there will be a spectrum of theories of this general form. These theories will vary in the way they explicate the notion of "enhancement": whether they construe this as involving increasing the probability of survival above a certain threshold, or simply increasing it significantly above what it would have been, and so on. We are not attempting to find and defend *the* correct propensity theory, but only arguing that a propensity theory offers the most promising theory of functions.

Fourthly, there is the question as to whether the scope of the propensity theory is limited to biological functions or whether it can be extended, in some sense, to artifacts.

Obviously, like the etiological theory, the propensity theory could be part of an overarching, disjunctive theory which analyses biological functions in terms of bestowing propensities for survival and the function of artifacts in terms of prior representations. We noted earlier some problems for a backward-looking theory, even for artifacts. Yet surely representations should have some causal role in the case of consciously produced artifacts.

We are attracted here to a general, overarching theory, but one that concentrates on the *propensity* for selection. So a character or structure has a certain function when it has a propensity for selection in virtue of that character or structure's having the relevant effects. In the case of biological functions, we have a propensity for survival in the natural habitat, and so in such a habitat natural selection is likely to be operative. In the case of artifacts, we have a selection process clearly involving representations. But the representations are those at the time of selection, at the time of bestowing the function: now, so to speak. They need not be blueprints that antedate the first appearance of the prototype. Thus we feel there is a sense in which all functions have a commonness of kind, whether they be of biological characters or of artifacts.

VII. COMPARISONS

We hope to have led our readers to appreciate the attractiveness of the propensity theory of biological functions (and of functions generally). We conclude by making some direct comparisons between the etiological theory and the propensity theory.

On most biological examples, the etiological theory and our propensity theory will yield identical verdicts. There are just two crucial sorts of case on which they part company.

One sort of case which distinguishes the theories is that of the *first* appearance of a character that bestows propensities conducive to survival.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

On our theory, the character already has a function, and by bad luck it might not survive, but with luck it may survive, and it may survive *because* it has a function.

On the etiological theory, in contrast, the character does not yet have a function. If it survives, it does not do so because it has a function; but, after time, if it has contributed to survival, the character will have a function.

We think our theory gives a more intuitively comfortable description of such cases, at least in most instances. But there are variants on this theme, on which our theory gives less comfortable results. Suppose a structure exists already and serves no purpose at all. Suppose then that the environment changes, and, as a result, the structure confers a propensity that is conducive to survival. Our theory tells us that we should say that the structure now has a function. Over all, this seems right, but there are cases where it seems counterintuitive. Consider, for instance, the case of heartbeats—that is, the *sound* emitted when the heart beats. In this century, the heartbeat has been used widely to diagnose various ailments; so it has come to be conducive to survival. The propensity theory deems the heartbeat to have the function of alerting doctors. That sounds wrong. The etiological theory says the heartbeat has no such function because it did not evolve for that reason. That sounds plausible.

And yet, we suggest, the reason we are reluctant to grant a function to the heartbeat is not that it lacks an evolutionary past of the required kind. Other characters may lack an evolutionary past, yet may happily invite attribution of a function. Rather, our reluctance to credit the heartbeat with a function stems from the fact that the sound of the heartbeat is an automatic, unavoidable by-product of the pumping action of the heart. And that pumping action serves other purposes. Although the heartbeat does (in some countries, recently) contribute to the survival of the individual, it does not contribute to survival of the character itself. The character—heartbeat—will be present in everyone, whether or not doctors take any notice of it. Although it "contributes" to survival, it is a redundant sort of contribution if it could not fail to be present whether it was making any contribution or not.

Perhaps the propensity theory should be carefully formulated in such a way as to rule out such "automatic" contributions to survival. Nevertheless, we will note only that, although the example of heartbeats seems initially to count against a propensity theory, there are other examples, and wider theoretical considerations, which count in its favor. Further, the case introduces many complications. For these reasons, it cannot be regarded as, in any sense, conclusive.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

So much for cases of *new* survival-enhancing characters. There is a dual for these cases: that of characters that were, but are no longer, survival enhancing. These cases, like the former cases, serve to distinguish between the etiological and propensity theories. If a character is no longer survival-enhancing (in the natural habitat), the propensity theory deems it to have no function. The etiological theory, in contrast, deems its function to be whatever it was that it used to be, and was evolved for.

In general, we think the propensity theory gives the better verdict in such cases. Under some formulations, our judgment may be swayed in favor of the etiological theory. We may be inclined to say that the function of a character is to do such and such, but unfortunately this is harmful to the creature these days. Yet surely the crucial fact is, really, that the function *was* to do such and such. It serves no pressing purpose to insist that its function still is to do that. Especially not, once we have passed the threshold over which we redefine the creature's natural habitat. If a character is no longer survival-enhancing, because of a sudden and recent change in environment, we may continue to refer its natural habitat to the past. Consequently, our propensity theory will continue to tie functions to what would be survival-enhancing in the past habitat. In such cases, there will be no conflict between the judgments of our theory and those of the etiological theory.

The test of examples and counterexamples is important. Yet in this case, in the analysis of functions, there is a risk that it will decay into the dull thud of conflicting intuitions. Similarly with intuitions as to how unified should be the analyses of functions for biological characters and for artifacts.

For this reason, we stress the importance of theoretical grounds for preferring the propensity theory. A propensity can play an explanatory causal role, whereas the fact that something has a certain historical origin does not, by itself, play much of an explanatory, causal role. Consequently, the propensity theory has a theoretical advantage, and this gives us a motive for seeking to explain away (or even overrule) apparent counterintuitions.

In a similar way, Darwinian evolutionary theory provides strong theoretical motives for analyzing fitness in a certain way. Our intuitions—our unreflective impulses to make judgments—have a role to play, but not an overriding one.

JOHN BIGELOW
ROBERT PARGETTER

La Trobe University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CONTENT BY COURTESY*

A fully naturalistic account of mind may be the next, perhaps the final, distinctive contribution of philosophy to human inquiry. Achievement of this goal requires, at a minimum, an account of how a wholly physical object in a physical world can be the subject of mental states. And since a science of the mind would seem to meet that requirement, many simply equate the naturalistic enterprise with the distinct task of providing a basis for a science of the mind.[1]

A constraint on any science of the mind, as currently conceived by many,[2] is that individuals in the same physical states must be assigned the same psychological states. Philosophers divide over whether or not a psychology that invokes states identified by content, such as beliefs and desires, can satisfy the constraint. Some are dubious about the scientific promise of a concept like *content*[3]; others (perhaps the majority) see intentional psychology as "the only game in town" and, not surprisingly, invest enormous and ingenious effort into formulating an appropriate concept of content. For if content cannot be made scientifically respectable, then beliefs, desires, and

* Alan Berger, Jerry Samet, David F. Austin, and William G. Lycan made very helpful comments on this work, an earlier draft of which was presented at the University of Vermont. This project has been supported by the National Endowment for the Humanities, the National Humanities Center, and the Faculty Research Fund of Middlebury College.

[1] The fact that both Donald Davidson and Ludwig Wittgenstein raise doubts about the prospects of a distinctive science of the mind in the course of their impeccably naturalistic investigations of mind shows that naturalism does not require "scientism." See, for example, Davidson, "Mental Events," in Lawrence Foster and J. W. Swanson, eds., *Experience and Theory* (Amherst: Massachusetts UP, 1970), pp. 79–101; "Thought and Talk," in Samuel Guttenplan, ed., *Mind and Language* (New York: Oxford, 1975), pp. 7–24; "Psychology as Philosophy" in his own *Essays on Actions and Events* (New York: Oxford, 1980), pp. 229–239. And Wittgenstein, *Philosophical Investigations*, 3rd ed., G. E. M. Anscombe, trans. (New York: Macmillan, 1968), and *Remarks on the Philosophy of Psychology*, vols. I and II, G. E. M. Anscombe and G. H. von Wright, eds. (Chicago: University Press, 1980).

[2] In my forthcoming book, *Saving Belief*, I argue against this constraint, which follows from a particular conception of science, a conception which is now almost universally, though sometimes tacitly, adopted by practitioners of cognitive science. (Cf. Jerry A. Fodor, "Individualism and Supervenience," ch. 2 of *Psychosemantics*, forthcoming.)

[3] For example, see Stephen Stich, *From Folk Psychology to Cognitive Science: The Case against Belief* (Cambridge, Mass.: MIT/Bradford, 1983); Paul M. Churchland, "Eliminative Materialism and Propositional Attitudes," this JOURNAL, LXXVIII, 2 (February 1981): 67–90; Patricia S. Churchland, "A Perspective on Mind-Brain Research," *ibid*, LXXVII, 4 (April 1980): 185–207.

0022-362X/87/8404/0197$01.70 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

so on cannot be invoked by a science that hopes to give nonintentional, nonsemantic conditions for mental states. In that case, many would conclude that the mental states that we enjoy are not, in fact, states of believing, desiring, or other putative states characterized by content. Thus, the task of formulating a concept of content suitable for science has a degree of urgency.

The requisite concept of content is called upon to do two things: it must allow content to play a semantic role in the determination of truth conditions of an individual's beliefs, and it must allow "contentful" states to figure in the etiology of an individual's behavior in a way that permits generalizations about content. A concept of content that satisfies the general constraint on theories of the mind will be such that molecule-for-molecule replicas will have the same contents. Thus, the external environment is relevant to content only to the extent that it impinges on sensory surfaces. Since the needed concept of content abstracts from the environment in this way, it has been called *narrow content*.

If the concept of content is both to play its semantic role in the determination of truth conditions of attitudes and to be adequate for incorporation into a scientific theory that explains behavior, it must satisfy two more specific constraints, one semantic and the other physical:

(S) (Narrow content + context) → Truth conditions of beliefs

and

(P) If C is a narrow content, and S has a belief with narrow content C, and S' is a molecule-for-molecule duplicate of S, then S' has a belief with narrow content C.

Any concept of narrow content that jointly satisfied both (S) and (P) would have this consequence: if molecular replicas were in the same contexts, their beliefs would have the same truth conditions. From the point of view under consideration, to avoid circularity, contexts must be fully specified nonintentionally and nonsemantically—roughly, without (even hidden) benefit of 'that'-clauses.

Informal motivation for these constraints is easy to find.[4] As an instance of the general constraint on a science of the mind, (P) may be thought of as an entry condition for any candidate for narrow content. Although molecular identity is not necessary for sameness of narrow content, it must be sufficient; otherwise, a taxonomy of

[4] Fodor discusses such motivation at length in unpublished manuscripts (for example, "Narrow Content and Meaning Holism"). Also see his "Cognitive Science and the Twin-Earth Problem," *Notre Dame Journal of Formal Logic*, XXIII, 2 (April 1982): 98–118. Also see "Individualism and Supervenience," *op. cit.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

states identified by narrow content would be useless for the kind of science of behavior under consideration.

(S) likewise seems natural. A major obstacle, brought out vividly in Twin-Earth cases, to the formulation of an appropriate concept of content is that truth conditions of beliefs (and satisfaction conditions of desires, etc.) seem to be determined in part by the believer's environment. Although at least one champion of content, Jerry Fodor, takes the Twin-Earth cases to be something of a red herring, he does take them to show (at most) that we must distinguish between broad contents and narrow contents. Broad contents must be relativized to contexts, and, when they are, they still determine truth conditions. Broad contents are functions from narrow contents and contexts to truth conditions.

The reductive strategy is to confine the intentional component of beliefs and other attitudes to narrow content; then, scientific respectability of intentional psychology is secured by the requirement that narrow content, whatever it turns out to be, must satisfy (P). A candidate for narrow content that fails to satisfy (P) is a nonstarter; a candidate for narrow content that fails to satisfy (S) is not a concept of content at all.

The picture is this: The truth conditions of a belief result from two vectors, the "internal" states of the believer and the "external" environment. On this assumption, the individual's contribution to her intentional states (narrow content) is distinguishable from the contribution due to her environment. Then, an adequate concept of narrow content would capture the contribution of the individual's internal states to her over-all states of believing, and so on. On this picture, the truth conditions of an individual's beliefs are the product of an "inner" semantic component (narrow context) and an "outer" nonsemantic component (context). (S) captures this requirement.

A number of possible candidates for narrow content clearly fail. For example, beliefs construed *de re* cannot be narrow contents, since they are not shared by molecular replicas. Anyway, according to Fodor, the beliefs that explain behavior are specified *de dicto,* but such beliefs are still "wider" than what is shared by molecular replicas. So, although they are to contribute to the truth conditions of *de dicto* beliefs, narrow contents must be construed still more narrowly than the content clauses of attributions of *de dicto* beliefs if they are to satisfy (P).[5] A promising new proposal has been made recently by

[5] For discussion of these and related issues, see Hilary Putnam, "The Meaning of 'Meaning'," in his *Mind, Language and Reality: Philosophical Papers,* vol. 2 (New York: Cambridge, 1975), pp. 215–271; Tyler Burge, "Individualism and the Mental," *Midwest Studies in Philosophy,* IV (1979): 73–122; and my "Just What

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Fodor, who aims to construe narrow content as mental symbols denoting phenomenologically accessible properties of distal objects.

Here I shall examine Fodor's proposal, the next in a line of proposals for a concept of narrow content that is supposed to satisfy both (S) and (P). After arguing that that specific proposal, based on an observation/nonobservation distinction, fails to satisfy (P) alone, I shall present a general argument to show that *no* concept of narrow content can satisfy both (P) and (S). Alternatively, nothing that satisfies (P) can make the requisite contribution to truth conditions of attitudes, and, hence, no concept that satisfies (P) is a genuinely semantic notion.

PHENOMENOLOGICAL ACCESSIBILITY

On Fodor's view, the bearers of semantic properties, in the first instance, are mental symbols in a language of thought; so having such-and-such narrow content will be a property of certain tokens of Mentalese. Fodor's basic idea is that narrow content is expressible in a vocabulary that denotes phenomenologically accessible properties of distal objects, where phenomenologically accessible properties are discernible by casual observation.[6] Then, the truth conditions of beliefs that explain behavior are a product of narrow content together with context.

The approach may be illustrated by showing how it applies seemingly in the right way to Twin-Earth cases. The cases are familiar enough: an Earthian and her counterpart on Twin Earth, where what is called "water" is not H_2O but XYZ, both think (or say), "Water is wet," and behave similarly toward the stuff; yet their utterances and beliefs have different truth conditions. How do we give a unified account of their similarities and differences? On Fodor's view, what they share is narrow content, expressed in a vocabulary denoting phenomenologically accessible properties of distal objects. The narrow content of both beliefs (Earthian's and Twin-Earthian's) may be expressed like this: The local stuff that is transparent, drinkable, sailable on, etc., is wet. On Earth, this belief gets evaluated with respect to H_2O; on Twin Earth, it is evaluated with respect to XYZ. So, as we may have hoped, the beliefs that both Earthian and Twin-Earthian would express by saying, "Water is wet," differ in truth

---

Do We Have in Mind?", *Midwest Studies in Philosophy*, x (1985): 25–48, and "A Farewell to Functionalism," *Philosophical Studies*, XLVIII, 1 (July 1985): 1–13.

[6] Fodor, "Narrow Content and Meaning Holism." Narrow concepts may be expressed by Ramsey formulas as well as by constants denoting phenomenologically accessible properties of distal objects. [Distal objects/events are contrasted with ("proximal") surface irritations of, say, skin or retina.]

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

conditions; yet the Earthian and Twin-Earthian both seem to entertain the same narrow contents. A tidy resolution of Twin-Earth cases is significant motivation for a view of propositional attitudes.

Moreover, if Fodor is correct that sensory and linguistic input systems are "modular," the output of such systems may be understood in observation terms. In that case, there would seem to be independent psychological confirmation of a principled distinction between properties that are phenomenologically accessible (or, equivalently for Fodor, observable) and others.[7]

Although the attempt to construe narrow content in terms of phenomenologically accessible properties of distal objects is well motivated, I believe that it ultimately founders on the constraint (P). Recall that Fodor emphasizes—with good reason, in my opinion—that the properties denoted by representations are not just phenomenal properties (such as would be exemplified by sense data), but are properties of physical objects in the environment. Fodor's answer to the question: In virtue of what does a representation denote what it does denote? is broadly causal. In a critical discussion of causal accounts of representation, he remarks, "The point of all of this, I emphasize, is *not* to argue against causal accounts of representation. I think, in fact, that something along the causal line is the best hope we have for saving intentionalist theorizing, both in psychology and in semantics."[8]

Fodor has argued at length that "the internal language must be rich enough to express the extension of any natural language predicate that can be learned."[9] So Mentalese in general is not restricted to denoting phenomenologically accessible properties. Nevertheless, it may be that Mentalese can aid the proponent of narrow content that is to satisfy (P). Perhaps Mentalese, unlike natural language, represents phenomenologically accessible properties in such a way that:

(R) For any type of Mentalese representation, R, such that R denotes a phenomenologically accessible property, if S_1 and S_2 are molecular duplicates, then S_1 has a token of R if and only if S_2 has a token of R.

[7] See Fodor, *The Modularity of Mind* (Cambridge, Mass.: MIT/Bradford, 1983), and "Observation Reconsidered," *Philosophy of Science*, LI, 1 (March 1984): 23–43. However, see also my forthcoming book for a discussion that shows that psychological considerations about modularity of input systems is idle as far as specifying a concept of narrow content is concerned.

[8] "Semantics, Wisconsin Style," *Synthese*, LIX, 3 (June 1984): 231–250, p. 234. Emphasis his.

[9] *The Language of Thought* (Cambridge, Mass.: Harvard, 1979), p. 82.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

For the moment, suppose that (R) is true. Then, it may seem that a concept of narrow content that satisfies (P) is at hand. The proposal would be that narrow contents should be construed as Mentalese representations of phenomenologically accessible properties of distal objects. Then, by (R), molecule-for-molecule duplicates have the same narrow content.

Unfortunately, (R), together with two other constituents of Fodor's view and the assumption that *being red* is a phenomenologically accessible property, leads to a contradiction. For it will emerge that the following triad is inconsistent:

(1) Mentalese representations denote properties of distal objects to which their tokens are causally connected.
(2) If two tokens of Mentalese representations denote different properties, then they are of different types of representation.
(3) (=R) For any type of Mentalese representation, R, such that R denotes a phenomenologically accessible property, if S_1 and S_2 are molecular duplicates, then S_1 has a token of R if and only if S_2 has a token of R.

To bring out the contradiction, I shall resort to a (regrettably exotic) thought experiment.

Consider another possible world; call it "Mars." Suppose that central to the Martians' religion (whose origins are lost in prehistory) is the wearing of special glasses. Imagine the elaborate "emplacement" ceremony practiced on infants near birth. The glasses are attached in such a way that they cannot be removed, except when exchanged for larger pairs as the youth matures. The actual exchanges, considered quite dangerous, are also attended by ritual, during which the young person is unconscious. Since attempts to remove the glasses (except during the exchange rituals) are taboo, no one has ever been known to remove the glasses.

Further suppose that, as it happens, there are no red objects on Mars; nothing on Mars reflects the appropriate wavelengths for red. Nevertheless, Martians do enjoy a full range of experiences, because their ritual glasses (little computers, actually) transform certain ambient energies that would be perceived as grey in the absence of the glasses, so that, when stimulated by light reflected from certain grey objects, the Martians' transducers are affected in exactly the same way that ours are when we see red.

Now suppose that there is an Earthly religion, practiced similarly. The only difference is that the glasses used by the Earthians are clear.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Now consider an Earthian and a Martian who are molecularly identical; the only differences between them are environmental: no red distal objects on Mars, but compensating red glasses; red distal objects on Earth, but clear glasses. Neither an individual's transducers nor visual input systems can detect whether the individual is in an Earthian or Martian environment.

Earthian and Martian are the same in their computational, functional, neurophysiological and other internal states. The relevant difference between the Earthian and her Martian counterpart is this: Most of the beliefs that the Earthian would express by uttering 'red' are about red things, but none of the beliefs that the Martian would express by uttering the acoustically identical sound are about red things; most are about grey things.

Now, suppose that the English-speaking Earthian, presented with a red ball, tokens the Mentalese representation that some round things are red, which in part causes her to utter in English, "Some round things are red." At the same time, the Martian is presented with a grey ball, which causes in her the same activity of transducers and input systems (and central processing, to whatever extent it figured in etiology of the Earthian's narrow content). The result of this mental activity is that the Martian tokens a Mentalese representation molecularly identical to the Earthian's. This representation in part causes the Martian to utter, in what sounds like English but is not, "Some round things are red."[10]

Now, by (1), the Earthian's and Martian's Mentalese representations denote properties of distal objects to which they are causally connected. The Earthian's relevant representations are caused in the most straightforward way by red things. The causal chain is traced from red objects that reflect light of a certain wavelength to stimulation of visual transducers by light of that wavelength, through the input systems, perhaps through some central processing; the output of these computations is a Mentalese representation, R_e. R_e denotes the phenomenologically accessible property of *being red*.

But, if it is by virtue of its causal connection to red things that the Earthian's Mentalese representation, R_e, denotes the property of being red, then the Martian's Mentalese representation, R_m, does not denote the property of being red. For no Mentalese representation of any Martian is causally connected in any way to any red thing;

[10] The Martians do not speak English because what sounds like 'red' in Martian is a different word from the English 'red'. Unlike 'I', 'here', and 'now', 'red' is not an indexical, which gets evaluated differently in different contexts.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

there are no red things on Mars. Indeed, the cause of the Martian's representation is traceable back to grey things, which reflect light of a certain wavelength, which is transformed by the glasses into light of the wavelength reflected by red things, and then impinges on her visual transducers in exactly the same way that the Earthian's transducers are stimulated. Moreover, this causal chain is not idiosyncratic; all over Mars, there is a systematic connection between grey things and Mentalese representations of the same type as R_m.

Now the contradiction is apparent: By (1) and (2), R_e and R_m are different types of representation, but, by (3), R_e and R_m are the same type. This vitiates the proposal that narrow content be construed in terms of Mentalese representations of phenomenologically accessible properties of distal objects.

The point can be made, specifically with regard to narrow content, from another direction. Assume that (1) and (2) are true. Now, what are the narrow contents of the belief expressed by the Earthian when she sincerely said in English, "Some round things are red," and the belief expressed by the Martian's emission of an acoustic twin? Since representations that constitute narrow contents are to represent phenomenologically accessible properties of distal objects, the narrow content of the Earthian's belief contains a representation of red, but the narrow content of the Martian's belief cannot be the same since the Martian has no representation of red. The narrow content of the Martian's belief contains a representation of fred, where 'fred' is an invented term denoting the representational property of those of the Martian's representations which are caused by grey things but which affect her transducers the way red things affect the Earthian's transducers. Since representations are typed by the properties of distal objects that cause them, a narrow content containing a representation of red is a different content from a narrow content containing a representation of fred. Hence, on the present proposal, our molecular duplicates fail to share their narrow contents, and (3) is false.

Or again: the extension of 'red' in English is the set of red things; the extension of 'red' in Martian is the set of grey things.[11] There-

[11] This seems in accord with Kripke's views as expressed in *Naming and Necessity* (Cambridge, Mass.: Harvard, 1980), fn 66, p. 128, and fn 71, p. 140. Kripke points out that "if we had different neural structures, if atmospheric conditions had been different, if we had been blind, and so on, then yellow objects would have done no such thing [as tending to produce certain characteristic sensations]." In that case, I take it, the extension of 'yellow', whose reference is rigidly fixed by "the external physical property that tends to produce in normal conditions such-and-such sensations," would have been different. Since different external physical properties

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

fore, 'red' in English and 'red' in Martian differ in extension. However, "for every predicate in the natural language it must be possible to express a coextensive predicate in the internal code."[12] So, 'red' in English and 'red' in Martian must have different Mentalese representations. So, it seems that recourse to Mentalese fails to show that molecular duplicates share narrow contents.

Let me dispose quickly of two objections: (i) Someone may object: Suppose that a child says, "Some animals are unicorns." We should say that she has a false belief about unicorns—even though there are no unicorns to have caused her representation of unicorns. Perhaps we should say likewise that the Martian had beliefs about red things, in which case Earthian and Martian may share narrow contents after all. But this objection is closed to someone who is trying to isolate an observation language (which is to exclude natural-kind terms). The advocate of an austere observation language would have a different story to tell about terms like 'unicorn' (or 'tiger') than about terms like 'red' in any case. Since 'unicorn' will not be an observation term anyway, beliefs about unicorns, like beliefs about water, will have different kinds of causal histories from beliefs about red things. So the fact that one may have beliefs about unicorns (in an intuitive sense of 'about') even if there are no unicorns to cause those beliefs is no grounds for supposing that one may have beliefs about red things even if there are no red things to cause those beliefs.[13]

(ii) Perhaps we should say that *if* the Martian's representations had had a typical or standard causal history, they would have been caused by red things. To this I should reply that we may assume that the Martians' case is atypical only in relation to a population that provides some norm of what is to count as typical.[14] Different kinds of causal chains are typical or standard in different possible worlds. Since there is no sense to an idea typical or standard causal chains across possible worlds, the typicality of the history of the Martian's representations can be determined only relative to her possible world, where she is undoubtedly typical. Since the representations of fred on Mars are typically (indeed, uniformly) caused by interaction with grey things that affect the Martians' transducers in the way that

---

satisfy the reference-fixing descriptions on Earth and on Mars, English 'red' and Martian 'red' have different extensions.

[12] Fodor, *The Language of Thought*, p. 85.

[13] Cf. Descartes, *Meditations* I.

[14] See David Lewis, "Mad Pain and Martian Pain," in Ned Block, ed., *Readings in the Philosophy of Psychology*, vol. I (Cambridge, Mass.: Harvard, 1980), pp. 216–222, for a population-relative version of functionalism.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

red things affect ours, it would seem question-begging to deny that the Martians are in normal circumstances. They are in circumstances normal on Mars.[15]

The only way that I see to avoid the conclusion that Earthian and Martian differ in narrow contents and, hence, that (R) is false or at least incompatible with other assumptions of Fodor's view, is to resort to phenomenalism, a desperate move that Fodor is understandably unwilling to make. But even that desperate move would be to no avail unless it could be shown that narrow contents in terms of properties of sense data could make the required contribution to truth conditions of beliefs. In any case, what has gone wrong, I think, is that, in order to avoid phenomenalism, Fodor invokes phenomenologically accessible properties of distal objects, i.e., properties of objects in the environment; but, if narrow contents represent properties of objects in the environment, then there is no guarantee that molecular duplicates share narrow contents.

To sum up: the account of narrow content in terms of phenomenologically accessible properties of distal objects was recommended for the way that it accommodates the Twin-Earth thought experiment. It turns out, however, that this account is heir to difficulties analogous to those raised by the Twin-Earth stories. The same difficulties that arise for natural-kind terms like 'water', in English or in Mentalese, also arise for observation terms like 'red', in English or in Mentalese.[16] So, it seems that narrow content understood as expressible in an observation language fails to satisfy (P).

SPEAKING ONE'S MIND

The hopelessness of narrow content satisfying (P) and (S) may be exhibited more generally. Not only does narrow content understood as mental representation of phenomenologically accessible properties fail to satisfy (P), but no genuinely semantic concept [such as would satisfy (S)] can satisfy (P). This is so because there can be molecule-for-molecule duplicates *with the same causal histories and in the same types of contexts*, described nonintentionally and nonsemantically, whose beliefs differ in truth conditions. In that case, if (S) is satisfied, then the duplicates' beliefs differ in narrow content and

[15] If one suspects that wearing glasses *per se* abrogates normal conditions (a suspicion, I imagine, confined to non-glasses-wearers), we could modify the example so that the normal atmosphere on Mars has the same effect as the glasses.

[16] Saul Kripke suggests certain similarities between natural-kind terms and terms like 'yellow'. See *Naming and Necessity* (Cambridge, Mass.: Harvard, 1970), note 66, p. 128.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(P) is violated; but if (P) is satisfied, then the duplicates' beliefs have the same narrow content and (S) is violated. A thought experiment will illustrate this possibility.

Suppose that two teenaged girls, both far from their respective homes, are invited to their first embassy party. Although their societies developed independently of each other, they are so similar in geography, climate, and so on, that the girls are molecular duplicates. One girl is from a country where standard American English is spoken, and the other is from a country where a language similar to English with a single exception is spoken. The exception is this: in English the word 'vodka' designates vodka, and in the mythical country the word 'vodka' designates gin. Assume that translators have established that the word in the remote language that sounds like the English word 'vodka' should be consistently translated as 'gin', because it occurs in contexts which are unproblematically translated into English as "*x* is what British officers in India drank with tonic in order to ward off malaria," "Hogarth depicted the tragic effects of excessive consumption of *x* in eighteenth-century England," and "Russia does not export *x*." Just as the Spanish word 'burro' means 'donkey' and the Italian word 'burro' means 'butter', so the English word 'vodka' means vodka and the mythical-language word 'vodka' means gin. In all other respects, translation between the mythical language and English is homophonic.

At the embassy party, one of the most popular refreshments is a clear liquid being served in shot glasses. Each girl gestures toward trays of drinks and utters what sounds like the English sentence, "Vodka is good to drink." In so uttering, one girl indicates in the mythical language that gin is good to drink, and the other girl indicates in English that vodka is good to drink. By producing the same "output," described nonsemantically, they say different things. There is no question about what each says.

Not only are the girls in the same current physical environment, but also they may have the same relevant histories, described nonintentionally. In particular, the episodes in which each acquires her beliefs may merit the same nonintentional descriptions. Suppose that the English speaker's beliefs about vodka have their origin in a training session for children of diplomats in which she is presented with a picture of a glass filled with a clear liquid and told that it is vodka and good to drink at receptions. Likewise, the non-English speaker's beliefs about gin have their origin in a training session in which she is presented with a picture of a glass filled with a clear liquid and told that it is gin and good to drink at receptions. The pictorial episodes

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

by which each girl acquires her attitudes may satisfy the same nonintentional descriptions; the pictures of the glass filled with a clear liquid may be indistinguishable. Both teachers may point to the pictures and emit the same sequence of sounds.

Thus, I think it plausible to conclude that, at the embassy party, the two girls are similar in all relevant nonintentional respects. They are in the same physical environment; they are in the same internal states; their relevant internal states are caused by the same physical environmental condition; they have the same relevant histories; they display the same linguistic and nonlinguistic behavior (nonintentionally described). Indeed, they are molecular duplicates in the same nonintentionally specified contexts.

We have seen that the girls' utterances, each in her own language, of what sounds like the English sentence, "Vodka is good to drink," have different truth conditions. Since each speaks sincerely and is a competent speaker of her native language, each expresses her belief —the English speaker that vodka is good to drink and the non-English speaker that gin is good to drink. These beliefs clearly have different truth conditions.[17]

Since the girls are in the same context, nonintentionally specified, and since their beliefs have different truth conditions, then, by (S), they differ in their narrow contents. But since they are molecular duplicates, then, by (P), they do not differ in their narrow contents. Therefore, (S) and (P) cannot both be satisfied.

The argument is perfectly general. Nothing hinges on whether or not the girls' beliefs are *de re* or *de dicto*. The point is independent of any particular specification of narrow content; I do not claim, for example, that "Vodka is good to drink" expresses narrow content. The claim is that, *however* narrow content is construed, according to (S), the girls differ in their narrow contents (since they are in the same contexts but their beliefs differ in truth conditions), but, according to (P), they do not differ in their narrow contents (since they are molecular duplicates).

(S) and (P) are not jointly satisfiable because they impose different identity conditions on narrow content—no matter how narrow con-

[17] Tyler Burge and Stephen Stich have deployed similar thought experiments. Unlike their thought experiments, however, the thought experiment here attributes no linguistic confusion or mistake on the part of either girl. Stich's goal is to discredit content; Burge's target is individualism. Burge also rejects "the view that the intentional can be accounted for in terms of the non-intentional," ("Other Bodies," note 5, p. 119). See also his "Individualism and Psychology," *Philosophical Review*, XCV, 1(January 1986): 3–45.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tent is specified. I conclude that (S) and (P) put incompatible demands on a concept of narrow content and that, therefore, no coherent concept can satisfy both constraints. The semantic and the nonsemantic [whether construed as physical, causal, functional, computational, syntactic, or any other states shared by molecular duplicates according to (P)] just do not pull in the same direction.

UNSPEAKABLE THOUGHTS

The suggestion of the argument just given is that the intentional and semantic, in principle, may be irreducible to the nonintentional and the nonsemantic.[18] Since that suggestion, which elsewhere[19] I pursue from other directions, flies in the face of the major thrust of the enterprise that has come to be known as Cognitive Science, I should defend the thought experiment against possible challenges. I anticipate two basic objections.

The first objection is that the girls really are not in the same context after all, and that the difference in context accounts for the difference in truth conditions of their beliefs. The other basic objection is that the girls really lack the beliefs ascribed to them, and that there is no difference between their beliefs correctly ascribed.

The first objection is that, although in many respects the girls are in the same context, in fact there is a difference in their contexts which accounts for the difference in the truth conditions of their beliefs. Now, how do the contexts differ? As the story was told, the girls are similarly located when they make their statements; they have similar causal histories (nonintentionally described); gin and vodka are both plentiful in their environments, past and present. The only differences between them are intentional or semantic. 'Vodka' denotes one drink in one language and another drink in a second language. One instructor intended to inform her pupil about vodka; the other instructor intended to inform her pupil about gin. It is even possible that the instructors themselves were molecular duplicates, who acquired their beliefs about gin and vodka, respectively, from other molecular duplicates. . . .

Suppose that these intentional chains eventually reach gin in the

[18] Both Stephen Stich and Paul Churchland entertain the possibility of irreducibility. The conclusions that I draw from irreducibility are substantially different from those that they draw, however. I discuss the differences in "Cognitive Suicide," forthcoming in *Contents of Thought: Proceedings of the 1985 Oberlin Colloquium in Philosophy*, Robert H. Grimm and Daniel D. Merrill, eds. (Tucson: Arizona UP, 1986).

[19] In "Just What Do We Have in Mind?," in "A Farewell to Functionalism," and in *Saving Belief*.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

one case and vodka in the other. Since the dubbings of gin and vodka may be indefinitely remote in space and time from the girls, why should these chains constitute a (nonintentional) difference in the *girls'* environments? All the girls' (nonintentionally specified) causal interactions are identical. Perhaps the objector would be willing to let the notion of environment or context balloon in such a way that indefinitely remote events may help define a context. At the least, such a move would preclude identifying contexts in terms of values for antecedently specifiable sets of parameters.

Moreover, such a move would be to no avail anyway unless there are causal chains, identifiable without intentional assumptions, leading from the dubbings to the girls. However, the way that a word is passed along from one member of the community to the next, at least on Saul Kripke's picture of names (*op. cit.*, 96/7), is that the learner *intends* to use it with the same extension as the teacher. Thus, even if we took dubbings to be nonintentional—and I have my doubts even about this—still, no one has shown how nonintentional specifications could be adequate to describe the appropriate chains leading to the girls. Of the countless causal chains leading to the girls' utterances, there is no obvious way to specify the relevant chains nonintentionally. So, on the plausible assumption that a (nonintentionally specified) context is determined at most by spatio-temporal proximity or (nonintentionally specified) causal connection, the girls do not seem to differ in environment.

The girls differ all right. And it is no feat to specify the difference semantically; but for purposes of determining truth conditions of attitudes [as in (S)], context cannot be specified in semantic or intentional terms, on pain of vicious circularity.

The second objection may be elaborated like this. Not only do the girls fail to distinguish between gin and vodka, but also neither is aware that there is a second liquid similar to the one about which she was told. Since the girls have exactly the same behavioral dispositions toward *both* gin and vodka, one may reasonably ask: Do the girls have enough information about the drinks to be said to have the concept of one drink but not of the other? If not, then, on one intuitive sense of 'about', they do not have beliefs about gin or beliefs about vodka; in that case, the example fails to show two molecularly similar individuals in the same circumstances nonintentionally described, but in different narrow intentional states.[20]

[20] Fred Feldman made a similar suggestion in conversation.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Since, however, we typically acquire beliefs in the fashion described and we typically lack omniscience, I see no credible way of denying that the girls have the beliefs ascribed. As the story was told, if either of the girls has the belief ascribed to her, the other has the belief ascribed to her. And, since the story may be filled in in indefinite detail, any principle that precluded both ascriptions would seem to be too strong. For it would seem to make the following a necessary condition of *S*'s believing that *F* is *G:* for any kind of thing that is *H* and not *F*, there are circumstances in which *S*, without acquiring new information, can distinguish things that are *F* from things that are *H*. (The murkiness of 'new information' does not affect the present point.)

Clearly, no such principle governs actual practice of ascription of belief. If it did, most of us would lack the belief that gold pieces are costly (since most of us are unable to distinguish gold from "fool's gold" without acquisition of new information), and most of us would lack the belief that U.S. hundred-dollar bills are green (since most of us are unable to distinguish them from counterfeits). Nor would adoption of such a principle seem desirable or even feasible; not only would it impoverish our store of beliefs, but also it would render miraculous our ability systematically to utter truths about, say, gold or hundred-dollar bills.

Thus, I think it unlikely that there is an adequate, principled way to deny that the girls have the beliefs ascribed. The English speaker believes that vodka is good to drink, and the non-English speaker believes that gin is good to drink. Each girl is simply unaware that there is a drink similar to the kind she has beliefs about. But there is no reason that this lack of further information should threaten the information that each has.

Moreover, if one wants to deny that the girls differ in belief, one will have to account for the difference in truth conditions of their sincere and competent utterances. For one said in English that vodka is good to drink, but, with the same kind of auditory emission, the other said in the mythical language that gin is good to drink. Anyone who claims that the beliefs they expressed thereby have the same truth conditions must conclude that the girls failed to express their beliefs.

Of course, from time to time, one fails to express one's belief in sincere utterance. For example, you might say, "The book on the table is well worth reading," if you believe that the book is Milan Kundera's latest. If Kundera's book has been replaced with one of Stephen King's, which you would not recommend, then you failed to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

express your belief. Or, you might make a linguistic mistake; wrongly thinking that 'pusillanimous' means "stingy," you might fail to express your belief if you said, "Scrooge is the most pusillanimous character in all of literature."[21] In both kinds of case, the speaker makes a mistake, and his failure to express his belief depends upon his having made that mistake. In the typical case, however, sincere and competent speakers succeed in expressing their beliefs.

If one tries to save narrow content by protesting that the girls really have the same belief, but fail to express it in their utterances, then one should give some indication, as I just did in the example above, of what prevents the girls from expressing their beliefs. Otherwise, the objector is just whistling in the dark.

Well, what would prevent them from expressing their beliefs? Nothing. Neither has made a mistake, linguistic or nonlinguistic. There is no further information on the basis of which either would be willing to admit that what she said was not really what she believed. Each is in excellent circumstances to say what she believed. However, I suppose that a friend of narrow content may just insist that such sincere and competent speakers of their language—whom no addition of new information, linguistic or nonlinguistic, would convince that they had "misspoken" themselves, or failed correctly to express what they actually believed—still did not express their beliefs.

The arbitrariness of such a "bite the bullet" strategy may be brought out by imagining a variation on the story. Suppose that there is no such country where what sounds like 'vodka' denotes gin, but that the course of the English speaker's life proceeds as before. Then, there are absolutely no grounds for denying that the English speaker expresses the belief that vodka is good to drink when she sincerely and competently says, "Vodka is good to drink." To insist in this case, with the non-English speaker out of the picture, that the English speaker fails to express the ascribed belief would be to impose an impossible standard on belief expression; any principle that precluded belief expression in the case of the English speaker considered alone would make it enormously difficult for us routinely to express our beliefs. Indeed, we may rightly wonder, if people fail to express their beliefs under the circumstances described, are we ever justified in supposing that anyone ever expresses her belief? And since there is no other belief that the English speaker would agree

[21] 'Pusillanimous' was a favorite source of examples for the late Herbert Heidelberger.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that she mistakenly expressed when she said, "Vodka is good to drink," if any belief helped cause her utterance, it is one unknown to and unknowable by the English speaker, and incommunicable to anyone else; it is an unspeakable belief. Recourse to such "beliefs" has the desperate ring of the *ad hoc*.

If belief/desire psychology is to be saved, it is a Pyrrhic victory, at best, to have to rely on unknowable and incommunicable beliefs. It is difficult to see how such beliefs could have explanatory efficacy. Indeed, it is difficult to see what would entitle such to be called "beliefs" at all. Thus, I think, insisting that the girls actually had the same belief, which they failed to express, is not a promising line of objection.

CONCLUSION

(S) and (P) are not jointly satisfied by any concept of narrow content. It does not matter whether the language of thought is a conventional language or not, nor does it matter whether narrow content is expressed in a restricted observation language or not. Here is one final way to put the dilemma: if an individual's narrow contents depend on the individual's environment, then (P) is violated, as we saw in considering observation terms in Mentalese. But, on the other hand, if an individual's narrow contents are independent of the environment, then it is a mystery how *narrow content* can be a semantic notion at all. That is, if (P) is satisfied, then, as we saw in the gin/vodka case, "narrow contents" (understood now as what is shared by molecular replicas) do not make the requisite contribution to the truth conditions of the individual's beliefs. So, a concept of narrow content that honors (P) is a concept of content only by courtesy.

LYNNE RUDDER BAKER

Middlebury College

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS AND CRITICISM

ATEMPORAL DURATION: A REPLY TO FITZGERALD*

PAUL FITZGERALD† thinks we "do not quite succeed" in our attempt** "to make the idea [of eternity] intelligible" (260). In his critical note he finds many things about eternity incoherent; most of them he simply dismisses without explanation. But he examines at some length one feature of the concept of eternity that is unquestionably essential to our account of it: the notion of atemporal duration.

Fitzgerald's Part I begins with a cursory review of our exposition of the Boethian definition of eternity as involving life, illimitability, duration, and atemporality; it concludes with his most pointed criticism of our work. His attempt to expose a contradiction in the definition depends on his subsequent presentation of three conditions, "the conditions we would expect to hold for *anything* that deserves to be called a 'duration' " (262). He characterizes these conditions more broadly as holding for *any* "*extensive* mode of being"—not only "temporal duration" but also "any spatial dimension" (*ibid.*). Fitzgerald's three conditions are: (1) "two distinct particulars can both have the kind of extension in question"; (2) "two particulars may both have the same or a different *amount* of the mode of extension in question"; (3) "by having different *positions* along the extensive dimension in question two qualitatively identical particulars can be numerically distinct" (262/3).[1]

In connection with condition 1 Fitzgerald says, "Perhaps two of God's thoughts or decisions can both have E-duration [the duration peculiar to eternity]. This feature of eternality may be impossible to reconcile with the divine simplicity. But that doctrine never had much to recommend it anyhow" (262). Fitzgerald dismisses the doctrine of simplicity without argument; we think the doctrine is an indispensable part of philosophical theology, as we have tried to show elsewhere.[2] And the concept of eternity we found in medieval philo-

* We are grateful to Calvin Normore for several helpful suggestions.

† "Stump and Kretzmann on Time and Eternity," this JOURNAL, LXXXII, 5 (May 1985): 260–269.

** "Eternity," this JOURNAL, LXXVIII, 8 (August 1981): 429–458; cited as "Eternity."

[1] Although condition (1) is superfluous, entailed by (2) and perhaps also by (3), we will consider the three conditions as Fitzgerald presents them.

[2] See our "Absolute Simplicity," *Faith and Philosophy*, II (1985): 353–382.

0022-362X/87/8404/0214$00.60 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sophers and undertook to defend is the concept of the mode of existence appropriate to a perfect, absolutely simple being; eternity as we see it must be compatible with simplicity. So, in constructing his objection to the concept of eternity around an interpretation of it that is incompatible with the doctrine of simplicity, Fitzgerald is mounting an attack that is sure to miss its target. Nothing that is incompatible with divine simplicity can count as E-duration.

For our purposes, it will be convenient to consider Fitzgerald's second and third conditions in reverse order. Like condition 1, condition 3 is incompatible with God's simplicity, especially as Fitzgerald illustrates it. But the supposition of the occurrence of two of God's mental acts at different points in E-duration (regardless of the sameness or difference of their contents) is even more clearly incompatible with the doctrine of eternity we presented. That is because the atemporality of E-duration is the direct consequence of eternity's having been defined as "the *complete* possession *all at once* of illimitable life" ("Eternity," 433/4). If the life that is God's eternal existence were to include acts with "different locations in E-duration," it would not be a life possessed completely all at once, a mode of existence atemporal in the only way eternity must be atemporal. Condition 3 cannot apply to E-duration in itself—not because E-duration is not an extensive mode of existence, but because it is (and was introduced by Fitzgerald as) an *atemporal* extensive mode of existence.

Fitzgerald's treatment of condition 2 develops complications that provide a transition to his direct criticism of our conception of atemporal duration. His first version of the theological illustration of (2) consists in the suggestion that "perhaps one of God's mental acts may have greater E-duration than another" (262). But when Fitzgerald applies condition 2 to E-duration, he applies it under a revised supposition: "even if every [divine] mental act as a whole has infinite E-duration, at least there must be different E-durational *subphases* of E-duration" (*ibid.*). This amounts to a revision of condition 2; we will call it (2a).

Condition 2a turns out to be essential to Fitzgerald's criticism of us. In its most general and least controversial form it might be expressed this way: (2a) In every extension, no matter what its mode of extension may be, there must be extended subphases having that same mode of extension. Consider the length of a continuous tabletop, one of Fitzgerald's nontheological illustrations of his second condition. Is it clear that in that length there must be lengths? Of course—infinitely many of them—as long as it is clearly understood that we are talking about divisions that *can* be made in it, physically

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

or conceptually, and not about already real constituents of it. And what is true of the application of (2a) to the length of the tabletop is likewise true of its application to beginningless, endless time—except that time's divisibility into years, months, and days is solely conceptual. But, in Fitzgerald's view, unless condition 2a is satisfied by E-duration as well, "the suggestion of quantitative limitlessness in 'infinite duration', 'remaining', 'enduring', and 'interminabilis' is just bogus. And we will not have distinguished eternality from bare timelessness, nor shown why the apt spatial analogue is the endless line rather than the unextended point" (262/3). So he thinks of (2a) as the condition critically necessary for the certification of E-duration as not only atemporal, but also durational.

According to Fitzgerald, both temporal duration and E-duration are actually, not just potentially and not just conceptually, divided into subphases, or composed of distinct actual parts. "I do not see how there can be infinitely extended duration with no finite parts. . . . Of course, given an Aristotelian doctrine that what is extended is not necessarily composed of actual parts but perhaps only of potential parts, one might suggest that God's E-duration is only potentially divisible into subphases. But that sits ill with any doctrine of God as Pure Act" (263/4). Neither Aristotle nor anyone else who takes time to be essentially continuous would characterize its divisibility by describing it as "composed . . . of potential parts." It is not made up of components at all, actual or potential; instead, it is *potentially divisible conceptually*. And Fitzgerald provides no reason for thinking that the subphases of E-duration are to be treated otherwise than those of temporal duration. As for the theological difficulty Fitzgerald alleges, *in his own nature* God is pure actuality, but nothing in that claim prohibits our conceiving of God as successively actualizing various potentialities, when we conceive of him from our temporal point of view. And, since the potential divisibility of any duration is conceptual only, there is no discrepancy between any possible divisibility of E-duration and God's nature as pure actuality.

Fitzgerald bases his direct criticism of us on his three conditions for duration, confronting us with a dilemma whose first horn is this: "If the analogy between E-duration and the more familiar modes of extension fails at any of these three points, one wants to know why, and whether the comparison then survives at all" (263). We have already noted several respects in which the analogy does fail at each of those three points, and we have said why it fails. Since we are nevertheless committed to the comparison between eternity and more familiar modes of extension, at least as regards maintaining

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that eternity is genuinely durational as well as genuinely atemporal, we think of ourselves as grasping this dilemma's first horn.

Here is its second horn: "If the analogies hold all the way, then one wants to know why E-duration isn't just another temporal dimension!" (263). Fitzgerald thinks this is the horn we choose: "Stump and Kretzmann save us from another, divine time only at the expense of deflating E-duration to an unextended point" (263). Such a move did not occur to us and would, in any case, have been confused, since the deflation Fitzgerald describes would entail the repudiation of the analogies in support of which it was supposed to have been undertaken. To confirm his view of our position, he cites a pair of passages in which we point out that eternity's atemporality entails that eternity involves "a duration without succession" (quoted 263), that for an eternal entity "there is no past or future, no earlier or later *within* its life; that is, the events constituting its life cannot be ordered sequentially *from the standpoint of eternity*" (quoted 263, from "Eternity," 434; second italics added by us here). We claimed, in other words, that the *successiveness* essential to temporal duration cannot be a feature of eternity in itself. And, as our use of the phrase "a *duration* without succession" shows, we do not take that claim to reduce eternity to a nonextensive mode of existence. As for successiveness in eternity *from a temporal point of view,* any conceptual convenience it might appear to offer is outweighed by the distortion it involves. Talking about the sun's setting (from a terrestrial point of view) is a harmless convention because the reality is so well understood, but talk about what God was doing before creation, for instance, is at best a misleading fiction.

Fitzgerald, however, interprets our denial of successiveness in eternity quite differently: "In other words, we do not really have an extensive mode in eternity at all, given not only that past and future and earlier and later are inapplicable, but that there is no analogue of them" (263) . . . "No, one cannot have literal infinite duration without ordered, finitely extended subphases" (264). Discounting Fitzgerald's realism regarding durational subphases, we would, of course, accept this latter judgment as applied to time. So what is it about eternity that warrants our describing it as literally durational even though devoid of "ordered, finitely extended subphases"?

We think we went some way toward answering that question in our article's section IV, "Atemporal Duration and Atemporal Life" (444–447), which, in our view, has not been squarely addressed by Fitzgerald's criticisms. But we will take this opportunity to supplement the material in that section in ways more directly addressed to Fitzgerald's concern.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The philosophers who discovered eternity saw it as the mode of existence requisite for whatever they saw as the foundation of temporal existence—for being as contrasted with becoming, for the absolutely perfect being as contrasted with the various and necessarily imperfect manifestations of being itself in material, temporal being. The heart of their conception of that mode of existence is the notion of atemporal duration. In the view of philosophical theologians such as Augustine, an analysis of time shows temporal duration to be only apparent duration, essentially inappropriate to the existence of God. The existence of a typical existent creature is spread over years of the past, through the present, and into years of the future. But neither its past nor its future is extant for any creature —the child you were, who appears in family snapshots, no longer exists—and any creature's present must be understood as a durationless instant at which its past is continuous with its future. Such radically evanescent existence could not be the existence of an absolutely perfect being. The mode of existence for such a being, the permanent, utterly immutable actuality on which the evanescent realm of becoming depends, must be characterized by genuine, paradigmatic duration. And such duration must be fully realized duration, none of which is already lost or not yet gained—an atemporal infinite duration, in short. So the classic ancient and medieval expositions of the concept of eternity are attempts to frame the notion of a mode of existence consisting wholly in a present that is limitless rather than instantaneous. Such a present is indivisible, like the temporal present, but it is atemporal in virtue of being limitless rather than instantaneous, and it is in that way infinitely enduring.

Some of the misunderstanding the concept of eternity seems especially vulnerable to is a consequence of the strain language undergoes when it is stretched to accommodate things utterly outside the experience it is founded on—a strain many of us learn to adjust to when the sciences require it of us. And the use in this context of terms such as 'duration' or 'the present', whose meaning is almost invariably expressed in terms of time, obviously invites trouble. In particular, many of Fitzgerald's criticisms of our account of eternity depend for their persuasiveness on his interpretations of claims entailed by the doctrine of eternity in terms of the ordinary temporal connotations of the key words in those claims. He objects, for instance, to our observation that there can be no succession in eternity, because he thinks that the denial of succession must "tear the very idea of duration . . . to shreds" (262). We are sympathetic to the claim that a denial of succession would tear the very idea of temporal duration to shreds, but, as our discussion of condition 2a indicates,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

atemporal duration, conceived as a beginningless and endless present, cannot admit of succession. We cannot even describe atemporal duration as divisible, because the familiar divisions of temporal duration give rise to the notion of a plurality of really distinct intervals and pick out one interval as earlier than or past relative to another that succeeds it. But neither successiveness nor plurality of intervals can be features of the atemporal duration that characterizes the mode of existence of the absolutely perfect being.

Finally, although spatial metaphors for eternity or time are troublesome even when they are helpful, it may be worth pointing out that both "the endless line" and "the unextended point"—often mentioned by Fitzgerald as if they were the only candidates for the role of an "apt spatial analogue" of eternity—strike us as providing more trouble than help. The closest we have been able to come to a picture that sometimes seems genuinely helpful (although it is still very crude and in many respects disanalogous) is this: two infinite parallel horizontal lines, the upper one of which, representing eternity, is, entirely and uniformly, a strip of light (where light represents the present), while the lower one, representing time, is dark everywhere except for a dot of light moving steadily along it. If the misleading representation of eternity as a line can be tolerated, the light, in the representations of eternity and of time, should be interpreted as an indivisible present. For any instant of time as that instant is present, the whole of eternity is present at once; the infinitely enduring, indivisible eternal present is simultaneous with each temporal instant as it is the present instant.

On the doctrine of eternity, atemporal duration is existence possessed completely, all at once, present entirely to its possessor, whereas temporal duration is existence possessed with radical incompleteness, a bit at a time, mostly already lost to its possessor or not yet possessed. Atemporal duration is the genuine, paradigmatic duration of which temporal duration is only the moving image. God's illimitable atemporal duration is the basis of all temporal duration, any instance of which is correctly called duration only analogically since it is only a partial manifestation of the paradigmatic, genuine duration which is "the complete possession all at once of illimitable life."

ELEONORE STUMP
Virginia Polytechnic Institute and State University

NORMAN KRETZMANN
Cornell University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

DOES "FITNESS" FIT THE FACTS?

A REPLY TO WILLIAMS AND ROSENBERG'S REJOINDER*

MY critical remarks[1] on Alexander Rosenberg's article on fitness[2] have elicited a rejoinder from Mary Williams and Rosenberg himself.[3] They charge that my criticisms are a "tissue of misunderstandings" (738); since they misunderstand my own position in fundamental ways, it may help to try to clarify the points that divide us. In the interest of brevity, I will ignore technical issues concerning the internal correctness of Williams's axiomatization[4] and will focus on questions of broader philosophical interest.

Williams and Rosenberg say that I "do not think there are any evolutionary laws" (745) and that I do not grant "the distinction between laws and initial conditions" (742). On the contrary, my article contends that the laws of evolution go beyond the meager handful that Rosenberg recognized. I also have no quarrel with the idea that laws must be "nomological generalizations" (739) and cannot refer to spatiotemporal particulars.[5]

In my view, evolutionary theory includes a wealth of lawlike generalizations; Williams and Rosenberg have excluded huge classes of these from the theory for no good reason. Evolutionary theory includes *source laws* and *consequence laws* for the various forces it recognizes as possible causes of evolution. A source law describes the circumstances that can bring a force into being; a consequence law describes the evolutionary changes that follow from the existence of a force or ensemble of such. There are several forces of evolution; they include more than just natural selection. This is my picture, which I think helps clarify how the theory of evolution is organized. For Williams and Rosenberg, evolutionary theory includes little more than the single generalization that heritable variation in fitness

* My thanks to the National Science Foundation for financial support.

[1] "Fact, Fiction, and Fitness," this JOURNAL, LXXXI, 7 (July 1984): 372–383.

[2] "Fitness," *ibid.*, LXXX, 8 (August 1983): 457–473.

[3] " 'Fitness' in Fact and Fiction," *ibid.*, LXXXII, 12 (December 1985): 738–748.

[4] "Deducing the Consequences of Evolution," *Journal of Theoretical Biology*, XXIX (1970): 343–385.

[5] Indeed, I begin *The Nature of Selection: Evolutionary Theory in Philosophical Focus* (Cambridge, Mass.: Bradford/MIT Press, 1984) by distinguishing historical hypotheses from lawlike generalizations in just this customary way.

0022-362X/87/8404/0220$00.50 © 1987 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

implies evolution. Why do they restrict the term 'evolutionary theory' to so narrow a compass?

Although Williams and Rosenberg talk interchangeably about "evolutionary theory" and "the Darwinian theory," I think they really have in mind the theory of natural selection. But, even if they wish to describe the theory of natural selection, why do they exclude generalizations about the Mendelian mechanism, on the one hand, and source laws, on the other?

They offer three reasons for thinking that population genetics is not part of evolutionary theory. They correctly point out that genes are not necessary for traits to be heritable, that evolutionary explanations can be sought for why the Mendelian mechanism is so pervasive, and that a Mendelian mechanism is compatible with non-Darwinian theories of evolution and even with no evolution at all (740/1).

None of these reasons comes close to showing why "evolutionary theory" should be restricted in the way Williams and Rosenberg demand. It is true that Mendelism is not necessary for heritability, but neither are many other things that are part of evolutionary theory. It is true that evolutionists may try to provide selectionist explanations for why the Mendelian mechanism is so widespread here on earth, but attending to the distinction between laws of nature and initial conditions will show that this is beside the point. The *generalizations* of population genetics are just that; one should not confuse them with the historical hypothesis that life on earth follows the Mendelian format. And, finally, it is true that the existence of Mendelism is compatible with non-Darwinian evolution and with no evolution at all; but why must each part of evolutionary theory imply that Darwinian evolution is what the living world is up to?

I offered R. A. Fisher's sex-ratio argument as an example of a source law that is part of the theory of natural selection. Fisher showed how, under suitable circumstances (qualitatively described, in Carnap's sense), the mix of sons and daughters a parent produces will determine its fitness. There are many such generalizations in evolutionary theory, each of them lawlike. Why do Williams and Rosenberg exclude them from "evolutionary theory"?

I found no reason for this in Rosenberg's original article; I still am in the dark after reading Williams and Rosenberg's reply. Source laws seem to be excluded because fitness cannot be type/type identified with any of its various "anatomical, physiological, behavioral, and environmental" (745) bases. I still don't understand why type/

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

type reducibility of fitness to the sorts of properties Williams and Rosenberg mention is necessary for source laws to be part of evolutionary theory.

Williams and Rosenberg also demand that source laws for fitness cite properties that are "mereologically, or part-whole-nomically connected to fitness, and which are supervenient on purely nonevolutionary properties of organisms" (745). It seems to me harmless to say that a parent's policy of dividing reproductive investment between sons and daughters is "part" of its fitness, as are numerous other properties that it has. These properties are "components of fitness" in the sense that selection for some and against others are component-selective forces, whose net effect is summarized in the assignment of a fitness level.

Williams and Rosenberg say that I attribute to Williams a view that she had explicitly repudiated—namely, that one could refute the tautology charge lodged against evolutionary theory by axiomatizing the theory with 'fitness' as a primitive expression (743). A few pages later, however, they take up my criticism that axiomatizing evolutionary theory in this way does not address the tautology problem (746). I had claimed, by way of analogy, that the charge that "$F = ma$" is a mere definition of force is not refuted simply by axiomatizing Newtonian theory with 'force' as a primitive. Williams and Rosenberg reply that "showing that either force or mass must be a primitive in the *best* axiomatization of Newtonian mechanics has been the most conclusive way to lay to rest the suspicion that the first law is vacuous" (746).[6]

I cannot see how to reconcile their claims of misinterpretation (743) with their subsequent discussion of what an axiomatization can hope to achieve (746); I therefore persist in believing that they hold that an axiomatization can show that a given generalization about fitness is not a tautology.

A theory can be axiomatized in various ways; which one we judge "best" is a function of our interests. Let us suppose that our interests in the theory of natural selection are sufficiently monolithic that there is one best axiomatization, and that 'fitness' is a primitive in it. I

[6] Williams and Rosenberg quote Max Jammer's *Concepts of Mass in Classical and Modern Physics* (Cambridge, Mass.: Harvard, 1961) in support of their optimistic view of the powers of axiomatization. The reader may wish to compare this evaluation with Lawrence Sklar's pessimism in *Space, Time, and Space-Time* (Berkeley: California UP, 1974), pp. 294–296.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

claim that this leaves entirely unsettled the question of whether propositions employing this concept are analytic or synthetic, a priori or a posteriori.

The axioms of a theory are usually chosen because they are, in some sense, fundamental. But even the most fundamental aspect of a theory can be thrown into doubt by observations, if there is a competing theory that makes that fundamental principle a bone of contention. Euclid's fifth postulate was as "axiomatic" as one might wish, but that doesn't settle whether it is nonempirical. My point is simply that the position of a statement within an axiomatization usually reflects our interests in the internal organization of the theory; it cannot anticipate all the various theoretical challenges that might be brought to bear against the theory, and so it cannot see in advance which of them are empirically testable and which are not. Saying that a proposition is an axiom doesn't settle these epistemological questions. Neither does saying that a term in a proposition is a primitive. Even if 'fitness' were a primitive, it still might be analytic or a priori that the fit tend to outsurvive and outreproduce the less fit, in which case all the old tautology conundrums would remain to be resolved.

ELLIOTT SOBER

University of Wisconsin/Madison

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

The Structure of Biological Science. ALEXANDER ROSENBERG. New York: Cambridge University Press, 1985. xi, 281 p. $37.50.*

Rosenberg's book is organized around the single question he considers central to the philosophy of biology: What is biology's relation to the physical sciences? According to Rosenberg the question has only two possible answers. Those who give the first answer he calls *autonomists*. Autonomists "hold that the really important aims of biology and the appropriate methods for attaining these aims are so different from those of the other sciences that biological theory and practice must remain permanently insulated from the distinctive methods and theories of physical science" (16). In contrast, *provincialists* hold "that at best biology is a province of physical science, a dependency that can advance only by applying the methods of physical science and, nowadays especially, the methods of physical and organic chemistry" (16). In chapter 2 Rosenberg tries to explicate this dispute, explain its importance, and sketch his strategy for resolving it. Clearly this is a metaphysical question, but presumably it is to be answered by recourse to our best physical and biological theories. But here a crucial ambiguity arises. Are we talking about properties of theories circa 1986, or are we talking about the properties these theories will "ultimately" have (whatever that means)? If it is the former then we are dealing with an interesting, nontrivial, and answerable question. (The answer is that at present biology is largely autonomous of the physical sciences.) If it is the latter then it is not obvious that the question is answerable. The latter question, unfortunately, is the focus of the book.

In quickly reviewing the history of this dispute, Rosenberg correctly points out that earlier versions of autonomism, namely vitalism and holism, were subjected to devastating empiricist criticisms. He also correctly points out that at present it is the provincialist who holds the epistemologically suspect metaphysical thesis; and so the same sort of criticisms that were directed at autonomism in the nineteenth and early twentieth centuries are, or can be, directed at provincialism now. It is odd that Rosenberg thinks we must choose one of the two positions; for there is obviously a third position, that of being agnostic with respect to the first two. Our forefathers, the logical positivists, would have adopted the agnostic position, and

* I am grateful to Richard Burian, Owen Flanagan, Norbert Hornstein, and David Sanford for their helpful comments on an earlier version of this review.

0022-362X/87/8404/0224$00.50 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

there is no reason why the failure of their verification theory of meaning should lead us to lower our epistemological standards.

Although I think we should reject the choice Rosenberg tries to force upon us, Rosenberg defends a version of provincialism that depends on the reducibility of biology to physics. The major arguments for this conclusion occur in chapters 4 and 5, and I shall devote the bulk of this review to examining them. In chapter 4, Rosenberg discusses molecular biology, in particular molecular genetics, and argues that Mendelian genetics is reducible to molecular genetics. Although Rosenberg's discussion of contemporary work in molecular biology is informative, the details of this work play no role in his argument. Indeed, the argument rests on two metaphysical theses: (1) the universe is finite; and (2) the universe is deterministic. (The latter thesis is improperly named. As Rosenberg is well aware, our basic physical theories are indeterministic; thesis 2 is basically an assumption of causal regularity.) The argument is this: Thesis 2 implies that identical molecular configurations will result in identical biological properties (e.g., the property of being a Mendelian gene for eye color). And although different molecular configurations may underlie the same biological property, thesis 1 implies that there are only finitely many such molecular configurations. Thus, Rosenberg argues, the biological property is equivalent to a finite disjunction of the molecular descriptions of these configurations. It is clear that at present we do not have a molecular theory that would supply these disjunctions. Furthermore, Rosenberg is willing to admit that in some cases the required theory would be so unmanageably complex that no human could ever construct, use, or know the theory.[1] Thus the reducing theory is unknown and unknowable; but that, Rosenberg claims, is only an instrumental or practical limitation, not an epistemological one (110)!

If this argument has consequences, they are ontological rather than methodological. Rosenberg does not claim otherwise. But the ontological consequences are unclear. If Rosenberg's reductionism has no effect on the shape of biological theories and if our ontological commitments derive from our best theories, then this version of reductionism puts few, if any, constraints on our ontology. (It does implicitly rule out vitalism.) Thus I am unpersuaded by Rosenberg's claim that his reductionist thesis is far from vacuous (117). I suspect that our disagreement stems from differences over what constitutes

[1] It is interesting to note that, in considering the autonomy of psychological explanation, Jerry Fodor takes the disjunctive nature of the reduction to be the single best argument for autonomy. See "Special Sciences," in *Representations* (Cambridge, Mass.: Bradford Books, MIT Press, 1981).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

an adequate basis for drawing ontological conclusions; unfortunately Rosenberg's views on this are not explicitly articulated and defended.

Even if we do accept the argument of chapter 4, Rosenberg sees that it does not suffice to show the reducibility of all of biology to the physical sciences. As his excellent discussion of the molecular structure of hemoglobin shows, evolutionary considerations impinge on molecular biology to make it fundamentally different from biochemistry. Thus, unless evolutionary biology can be reduced to the physical sciences, the autonomist is proved right (119). Chapter 5 provides the main argument that evolutionary biology is so reducible. The argument depends heavily on an axiomatization of the theory of natural selection due to Mary Williams. Indeed, it depends on a misunderstanding of this axiomatization. Rosenberg argues that Williams's axiomatization has the sort of universality characteristic of theories in the physical sciences and, in particular, that it is true of any biological entities that satisfy the parent-of relation. He says, "If there ever were, or are now, or ever will be biological entities that satisfy the parent-of relation, anywhere in the universe, then they will evolve in accordance with this theory (or else the theory is false)" (152). He is mistaken.

Space does not permit a detailed presentation of the axiom system, but I think we can see Rosenberg's error by looking at axiom 5:

> Axiom 5. In every generation m of a Darwinian subclan D which is not on the verge of extinction, there is a subcland D_1 such that: D_1 is superior to the rest of D for long enough to ensure that D_1 will increase relative to D; and as long as D_1 is not fixed in D it retains sufficient superiority to ensure further increases relative to D.[2]

Axiom 5 is true of Darwinian subclans. What is a Darwinian subclan? Informally, a Darwinian subclan is a biological population all of whose members experience common selective forces. Formally, the term is undefined in Williams's axiomatization. Thus a Darwinian subclan is anything that satisfies the axioms, in particular axiom 5. So to say that axiom 5 is true of Darwinian subclans is just to say that it is true of what it is true of. The empirical question remains whether it can be applied to any particular population. It is easy to find biological populations satisfying the parent-of relation to which it does not apply. For instance, it does not apply to a laboratory population of genetically identical mice. Lacking variation, they are not subject to

[2] M. B. Williams, "Deducing the Consequences of Evolution: A Mathematical Model," *Journal of Theoretical Biology*, XXIX (1970): 366. In the book Rosenberg mistakenly substitutes 'Darwinian subcland' for 'Darwinian subclan' (144).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

selection. Perhaps more interestingly, it does not apply to populations of selectively neutral segments of DNA.[3] Contrary to Rosenberg, this does not mean that the theory is false. It just means that there are limits to its applicability. Unfortunately, Rosenberg thinks that the axiomatization applies to all populations of biological entities (see 138, 152, 168, 178). Whether this leads to or results from a fundamental misunderstanding of evolutionary biology I do not know, but it is associated with some of the most outrageous statements in the book. I shall mention just one. In applying the axiomatization to predator-prey interactions, Rosenberg blithely assumes that populations consisting of predator and prey species are Darwinian subclans. So, by deduction from Axiom 5, he claims that we can *know* that in all such predator-prey interactions there is an equilibrium value of the proportion of predators to prey (215). In fact, however, equilibrium values do not exist for all predator-prey systems. Again, this does not falsify Axiom 5. It just reflects the fact that real populations consisting of predator and prey species are not Darwinian subclans.

In spite of the above criticisms, I think Rosenberg's book has considerable merit. Although the book is organized around the autonomism versus provincialism issue, the chapters are fairly self-contained. For instance, one could profitably read chapter 7 on the species question without having read the earlier chapters. Also, each chapter contains a guide to the literature. Finally, the book is original even if the major theses are mistaken.

ROBERT N. BRANDON

Duke University

Identity, Cause, and Mind. SYDNEY SHOEMAKER. New York: Cambridge University Press, 1984. 363 p. Cloth $52.50, paper $15.95.

The fifteen essays collected in this volume span the period from 1968 to 1981. They have all been printed before, though Shoemaker has made a number of minor revisions and added two new postscripts. The topics treated include: self-reference, personal identity, memory, time and change, "grue," other minds, embodiment, dualism, phenomenal similarity, properties and causality, identity through

[3] See M. Kimura, *The Neutral Theory of Molecular Evolution* (New York: Cambridge, 1983).

0022-362X/87/8404/0227$00.60 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

time, functionalism. Shoemaker's discussion of these topics is marked by controlled rigor, argumentative density, honesty, caution, resourcefulness—and a degree of unrelenting heaviness. The essays are thus a model of what professional philosophy should be. I shall restrict my comments to two central themes in Shoemaker's thinking: the nature of properties, and the defensibility of functionalism. There is much else in the volume that might have been discussed.

Shoemaker claims that "the identity of a property is determined by its causal potentialities, the contribution it is capable of making to the causal powers of the things that have it" (213). This claim (CTP) is elaborately and convincingly defended, and its anti-Humean consequences are unflinchingly acknowledged. CTP has, as Shoemaker notes, been independently advocated by other philosophers (Achinstein, Armstrong, and others), but I think his development of the idea is the most sophisticated going. What he does not note, however, is the similarity between CTP and the Davidson-Nagel criterion of event identity: events are the same iff they have the same causes and effects. This criterion was originally proposed for event tokens, but it is readily generalized to event types: event types are the same iff they are apt to be caused by events of the same type and to cause events of the same type; and, when it is so generalized, it becomes virtually indistinguishable from CTP (event types being identified in terms of properties). And the two individuation conditions have this also in common: they are both circular. Causes and effects of events are themselves events, and the condition speaks of their identity; and the causal powers of a property have to be understood in terms of the instantiation of other properties—viz., those properties upon which the power is conditional and those properties in the instantiation of which the exercise of the power consists. But, as Shoemaker rightly says, such circularity is a problem only if one is bent on achieving a *reductionist* account of the notion of property (or event); instead, one can offer CTP as a mapping of the internal and reciprocal relations between the members of a family of notions— property, cause, power, change, etc. I think this is fair enough, but it ought to be noted that CTP, offered as an identity criterion, would not serve to pacify Quinean hostility to the notion of property; for Quine wants to be told when properties are to be counted the same *without* reliance upon a prior understanding of property identity (*mutatis mutandis* for events). The issue then is whether Quine is entitled to demand some external point of purchase on the family of notions in question. I would myself say not, and so it seems would Shoemaker, but it is worth noting the polemical limitations of the causal account.

I am a bit unsure about Shoemaker's official motivation for CTP: this is that only if CTP is true can it be explained how properties are

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

capable of engaging our knowledge, and our language, in the way they do" (214). My doubts have two sources. First, Shoemaker is here presupposing a causal theory of knowledge, justification, and reference, which is controversial in general and which appears clearly wrong in certain areas, e.g., mathematics. The fate of CTP should not, I think, be made to depend upon the fate of these other causal theories; and the unavailability of a causal account of the properties of numbers does not seem to prevent us knowing about such properties and referring to them. Secondly, it seems to me undesirable to discoverable real essences; but, he argues, this thesis appears implautions: surely we ought to be able to say something about the nature of properties, quite independently of our epistemic relation to them, which makes CTP an attractive theory of properties. (Note that Davidson's account of events is not epistemologically motivated.) Intrinsic plausibility is one merit of the account which ought not to be sneezed at; and the avoidance of some of the absurder consequences of runaway Humeanism is another point to press—we really *don't* want to accept the real possibility of a world in which precisely the same substances, properties, and events exist as in the actual world but the causal laws and relations are as different from those obtaining in the actual world as you like. It is not, of course, that Shoemaker is blind to these points—far from it—but I think he puts the emphasis in somewhat the wrong place.

Causality and properties figure crucially in Shoemaker's ingenious discussion of identity through time. Here the core idea is that persistence is a matter of the causal relations between successive property-stages of the continuant in question. Circularity objections loom over this thesis too, as Shoemaker quickly points out. These arise at two distinct points. There is, first, the problem of specifying *which* sorts of causal connection constitute persistence—the danger being that these must be defined as those which obtain *in a persisting object*. And, second, the appeal to properties involves us, via CTP, in the idea of powers, and powers are exercised by persisting objects over time. In response to this kind of difficulty for a reductive account of cross-temporal identity, Shoemaker settles for a supervenience thesis: "if in any possible world a set *S* of property instantiations makes up the history of a single persisting thing, then in any world that is qualitatively identical to that world, in the sense that the same properties are instantiated at the same space-time points, and also identical to it with respect to the sorts of causal connexion that hold between the various property instantiations, the set of property instantiations in that world which is the counterpart of *S* will likewise make up the history of a single persisting thing" (247). Supervenience theses are not in general trivial, but it remains unclear

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

whether this particular one can avoid a pretty malign kind of circularity, since the supervenience base appears to be tacitly understood as already including the allegedly supervenient fact.

Contrast this case with that of the supervenience of the mental on the physical: here we clearly don't need to rely upon mental notions to characterize the physical supervenience base. This does not of course make Shoemaker's account completely vacuous—he is still forging conceptual links among identity, cause, and property—but it shows how far short the account falls of the old ambition to tell us in what identity through time *consists*.

These reflections on persistence in general lead Shoemaker on to some conclusions about the persistence of persons. Philosophers of personal identity tend to approach their subject in isolation from the general issue of continuant persistence; it is good that Shoemaker rectifies this tendency. And there is a definite payoff in that we are made aware of what kinds of theory of personal persistence we can legitimately expect. A causal account, for example, where the causal relata are successive mental states, will be unavoidably circular and nonreductive; for it cannot take the notion of a mental state as prior to the notion of a person, and it is unclear that it can provide an independent account of the kinds of causal relations that determine personal identity. We also see, incidentally, that the standard opposition between reductionism about personal identity and outright conceptual primitivism about the notion is too crude.

In "Some Varieties of Functionalism" Shoemaker connects CTP with functionalism: if CTP is true generally, then it is true of mental properties, and CTP about mental properties is tantamount to a kind of functionalism about them since it individuates properties by their causal features. One can react to this subsumption point in two ways. One can take it as reinforcing functionalism about mental properties, or one can take it as trivializing that theory. Shoemaker takes it the first way; I am inclined to take it the second way. I see it as a potentially trivializing entailment because functionalism is no longer telling us what is *distinctive* of the mental—*every* property turns out to be a functional property. Even the holism of the mental, which functionalism advertised itself as capturing, turns out to be a special case of the general truth that properties have their causal powers conditionally upon which other properties they are coinstantiated with (a knife-shaped thing has the power to cut bread conditionally upon being made of steel, a particular desire has the power to produce a piece of behavior conditionally upon the agent's also having certain beliefs). What is not trivial, of course, is precisely which version of functionalism to adopt, i.e., which of the causal features of a mental property should be taken to be definitive of it. But that

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

mental properties are individuated by their causal role comes out as a straightforward consequence of the fact that *no* property can be pulled apart from its causal role. Seeing functionalism in this light also raises a suspicion about whether it has its conceptual priorities right: objects have the causal powers they have *because* of the properties they possess; they don't have those properties in *virtue* of having such powers. The asymmetry thus goes the wrong way to underpin a genuine reduction.

Shoemaker operates with a pair of distinctions which together yield four possible brands of functionalism. There is the distinction between analytical folk-psychology-based functionalism and empirical scientific-psychology-based functionalism; and there is the distinction between weak and strong functionalism—the weak kind defining mental properties by ineliminable reference to other mental properties, the strong kind totally eliminating such reference from its eventual definiens. These distinctions cannot be overstressed; the various theories they generate have very different consequences and prospects of success.

To see the significance of the weak/strong distinction, consider the following pair of theses: (a) physical properties can be defined in terms of their causes and effects, including their physical causes and effects; (b) physical properties can be defined in terms of their *mental* causal powers, i.e., their capacity to cause experiences in perceivers (phenomenalism). Weak functionalism allows concepts from the same domain as the definiendum concepts to appear in the definiens; strong functionalism (like phenomenalism) allows only concepts drawn from a different domain in the definiens. Success in the weak enterprise thus offers no encouragement at all to the strong project.

The distinction between analytical and empirical functionalism is important for a different reason, namely, the different conceptions of the semantics of mental terms implicit in the two theories. As Shoemaker rightly says, the empirical version seems committed to the thesis that mental terms denote natural kinds with scientifically discoverable real essences; but, he argues, this thesis appears implausible when we consider what we would say of possible creatures whose "depth psychology" differs from ours: we would *not* say that they could not have the same mental properties we have. Analytical functionalism, by contrast, implies just the opposite about mental terms: unlike terms for natural kinds, terms for mental properties *can* be defined analytically and a priori. All this seems to me quite right (indeed I have argued much the same thing[1]). The variety of

[1] In "Mental States, Natural Kinds and Psychophysical Laws," *Proceedings of the Aristotelian Society*, suppl. vol. LII (1978): 195–220.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

functionalism Shoemaker himself undertakes to defend is typically weak analytical functionalism (though he declares his adherence to the strong version). This, as he remarks, is closely akin to the old "conceptual-connection" thesis, whereas empirical functionalism is tantamount to the denial of that thesis. So different are the two theories that calling them both "functionalism" is more misleading than illuminating, in my view.

Qualia bulk large in a number of the essays: can functionalism give a satisfactory account of them? The weak/strong distinction is crucial to Shoemaker's reply to this question. He holds that if we help ourselves to the subject's introspective beliefs we can evade the absent-qualia objection (the same can be said of the inverted-qualia objection); for it is implausible to suppose that someone could have beliefs purportedly about his/her own qualitative states and yet not have these states. Of course, it may be that these introspective beliefs are not themselves functionally definable; but (i) this doesn't count against weak functionalism, and (ii) the reply shows (according to Shoemaker) that qualia *as such* are no special problem for a functionalist, weak or strong. I have two main worries about Shoemaker's argument here. The first is that appealing to qualitative beliefs in the definition of qualia does not help when we are dealing with creatures that have qualia but do not have introspective beliefs (presumably many animals are in this case). But then what is to show that they cannot have (weak) functionally equivalent counterparts that lack qualia? Moreover, since their qualia could be the same type of psychological state as ours, despite the difference in the beliefs we can respectively have, it would seem that qualitative beliefs cannot be *essential* to the functional definition of qualia. The second worry concerns Shoemaker's claim that his argument, if successful, shows that qualia pose no *special* problem for functionalism. It seems to me that this is so only on the assumption that the functional indefinability of qualitative beliefs, should they prove indefinable, does not depend upon their being specifically beliefs *about qualia*. For, if the *explanation* of their not being functionally definable is that they have a qualitative subject matter, and other sorts of beliefs *are* functionally definable, then it seems correct to attribute the failure of strong functionalism to the existence of qualia. Until it is shown that this is not so, Shoemaker's conclusion seems to me to be premature.

This is a stimulating collection of papers by an important philosopher. Anyone with a taste for tough abstract argument and rigorous formulation will come away from reading this volume feeling adequately sated.

Oxford University COLIN MCGINN
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

The editors report with sorrow the death of Morris Lazerowitz, Professor Emeritus at Smith College, on February 25, 1987, at his home in Northampton. He and his wife, Alice Ambrose, had taught at Smith College for thirty-five years, beginning in 1938; in their last nine years before retirement they jointly held the Sophid and Austen Smith Chair. Professor Lazerowitz had written extensively about logic, metaphilosophy (his coinage), and the philosophy of language, much of his writing coauthored by Professor Ambrose. At the time of his death he was 79 years old.

The editors report with regret the death of Gerald C. MacCallum on January 14, 1987, in Fairway, Kansas. Professor MacCallum joined the faculty of the University of Wisconsin in 1961, serving as Department Chairman from 1970 to 1972, and as Secretary-Treasurer of the Western (now Central) Division of the American Philosophical Association, 1967–69. He had recently completed *Political Philosophy,* published in 1986 by Prentice-Hall's Foundations of Philosophy series. Professor MacCallum was 61 years old at the time of his death.

Studies in Moral Philosophy is pleased to announce the establishment of a new book series, to be published by Peter Lang Publishing. The series will be edited by John Kekes, who invites proposals for original book-length manuscripts examining the nature of good lives, character and its development, virtues and vices. Manuscripts considering these topics in relation to literary works are especially welcome. Inquiries and submissions may be directed to: John Kekes, Department of Philosophy, SUNY/Albany, N.Y. 12222.

Student Pugwash USA announces its Fifth Biennial Student Pugwash USA International Conference, to be held June 28–July 4, 1987 at Stanford University. The theme of the Conference is "Choices for Our Generation: Ethics and Values at the Cutting Edge of Technology." The conference is limited to 100 students, each admitted to one of six working groups: Dilemmas for the Future of Computing; Roles for the Biotechnologies in International Development; Nuclear Proliferation and the International Control of Atomic Energy; Water—Politics, Pollution, and Supply; Reproductive Technologies; and Science and Technology in the Media. Each working group also includes several Senior Participants, leading professionals involved with the respective topics. Further information may be obtained from: Benjamin Austin, Conference Director, Student Pugwash USA, 505-B Second Street N.E., Washington, D.C. 20002.

Loyola University of Chicago and Loyola University Press are pleased to announce a new book series on Values and Ethics. The editors welcome proposals from scholars working on ethical issues related to the humanities, sciences, law, medicine, and the helping professions. Prospective authors should submit a letter of inquiry and summary outline indicating the manuscript's general area, specific topic or theme, contents, and length. A short

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

vita may also be enclosed. Proposals may be sent to: Gerald Gutek, University Press Series, Loyola University of Chicago, Graduate School, Chicago, Ill. 60611.

The Department of Philosophy at the University of Minnesota is pleased to announce that Arthur Caplan will assume the directorship of the Center for Biomedical Ethics, newly established in the Medical School at the University of Minnesota, effective June 1, 1987. Professor Caplan also holds a joint appointment in Surgery and in the Philosophy Department.

Williams College is pleased to announce that its first John William Miller Fellowship has been awarded to Assistant Professor Stephen Tyman (Southern Illinois), whose study is tentatively entitled *Affirmative Self-Control in the Ethics of John William Miller*. The Fellowship supports research for book-length studies on the work of John William Miller. Moreover, an annual prize is awarded for the best essay published in a recognized journal on some aspect of Professor Miller's work. The 1986 winner is Robert S. Corrington (Pennsylvania State) for his essay, "John William Miller and the Ontology of the Midworld," published in *Transactions of the Charles S. Peirce Society* (Spring 1986). Further information and applications for both the fellowship and the essay prize may be directed to: John William Miller Fellowship Fund, c/o The Library, Williams College, Williamstown, Mass. 02167.

NEW BOOKS: ANTHOLOGIES

COLODNY, ROBERT G., ed.: *From Quarks to Quasars: Philosophical Problems of Modern Physics*. Pittsburgh: University Press, 1986. xvi, 389 p. $55.

COOPER, DAVID E., ed.: *Education, Values and Mind: Essays for R. S. Peters*. Boston: Routledge & Kegan Paul, 1986. vi, 222 p. $26.95.

DEGRÉ, GERARD, and CYRIL H. LEVITT, eds.: *The Social Compulsions of Ideas*. New Brunswick, N.J.: Transaction Books, 1985. xiii, 264 p. $39.95.

DEMARCO, J. P., and R. M. FOX, eds.: *New Directions in Ethics*. Boston: Routledge & Kegan Paul, 1986. xi, 335 p. Cloth $29.95, paper $14.95.

DILMAN, ILHAM, ed.: *Philosophy and Life: Essays on John Wisdom*. Boston: Nijhoff, 1985. 345 p. $56.

DURANT, JOHN, ed.: *Darwinism and Divinity: Essays on Evolution and Religious Belief*. New York: Basil Blackwell, 1985. 210 p. $24.95.

EARLE, WILLIAM. *Festschrift*. See Casey.

ELSTER, JOHN, and AANUND HYLLAND, eds.: *Foundations of Social Choice Theory*. New York: Cambridge, 1986. 250 p. $32.50.

EZORSKY, GERTRUDE, ed.: *Moral Rights in the Workplace*. Albany: SUNY Press, 1986. xii, 312 p. Cloth $34.50, paper $10.95.

FISCHER, JOHN MARTIN, ed.: *Moral Responsibility*. Ithaca, N.Y.: Cornell, 1986. 260 p. Cloth $35, paper $10.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

FINDLAY, J. N. *Festschrift*. See Cohen.

FISKE, DONALD W., and RICHARD A. SHWEDER, eds.: *Metatheory in Social Science: Pluralisms and Subjectivities*. Chicago: University Press, 1986. x, 390 p. Cloth $35, paper $16.95.

FORESTER, JOHN, ed.: *Critical Theory and Public Life*. Cambridge, Mass.: MIT Press, 1985. xix, 331 p. $30.

FORTENBAUGH, WILLIAM W., PAMELA M. HUBY, and ANTHONY M. LONG, eds.: *Theophrastus of Eresus: On His Life and Work*, vol. II. New Brunswick, N.J.: Transaction Books, 1985. ix, 355 p. $49.95.

FOSTER, JOHN, and HOWARD ROBINSON, eds.: *Essays on Berkeley*. New York: Oxford, 1985. 256 p. $34.50.

FULLINWIDER, ROBERT K., and CLAUDIA MILLS, eds.: *The Moral Foundations of Civil Rights*. Totowa, N.J.: Rowman & Littlefield, 1986. vii, 206 p. Cloth $31.95, paper $14.95.

GOLDEN, JAMES L., and JOSEPH J. PILOTTA, eds: *Practical Reasoning in Human Affairs*. Boston: D. Reidel, 1986. x, 404 p. $79.

GRANDY, RICHARD E., and RICHARD WARNER, eds.: *Philosophical Grounds of Rationality*. New York: Oxford, 1986. 500 p. $29.95.

GRIFFIN, DAVID R., ed.: *Physics and the Ultimate Significance of Time: Bohm, Prigogine, and Process Philosophy*. Albany: SUNY Press, 1986. xv, 322 p. Cloth $44.50, paper $14.95.

HAAPARANTA, LEILA, and JAAKKO HINTIKKA, eds.: *Frege Synthesized*. Boston: D. Reidel, 1986. vi, 395 p. $69.

HAUSMAN, DANIEL M., ed.: *The Philosophy of Economics*. New York: Cambridge, 1985. viii, 415 p. Cloth $39.50, paper $14.95.

HELLER, AGNES, and FERENC FEHÉR: *Reconstructing Aesthetics*. New York: Basil Blackwell, 1986. vii, 255 p. $34.95.

HELLER, THOMAS C., *et al.*: *Reconstructing Individualism: Autonomy, Individuality, and the Self in Western Thought*. Stanford, Calif.: University Press, 1986. xiv, 365. Cloth $39.50, paper $11.95.

HICK, JOHN, and HASAN ASKARI: *The Experience of Religious Diversity*. Brookfield, Vt.: Gower, 1985. vi, 236 p. $39.95.

HOLLINGER, ROBERT, ed.: *Hermeneutics and Praxis*. Notre Dame, Ind.: University Press, 1985. xx, 296 p. Cloth $29.95, paper $12.95.

HORTON, JOHN, and SUSAN MENDUS, eds.: *Aspects of Toleration: Philosophical Studies*. New York: Methuen, 1985. 180 p. $29.95.

IHDE, DON, and HUGH J. SILVERMAN, eds.: *Descriptions: Selected Studies in Phenomenology and Existential Philosophy*. Albany: SUNY Press, 1985. xv, 300 p. Cloth $44.50, paper $19.95.

IHDE, DON, and HUGH J. SILVERMAN, eds.: *Hermeneutics and Deconstruction*. Albany: SUNY Press, 1985. xi, 309 p. Cloth $44.50, paper $19.95.

KENNINGTON, RICHARD, ed.: *The Philosophy of Immanuel Kant*. Washington, D.C.: Catholic University of America Press, 1985. 256 p. $31.95.

KENNY, ANTHONY, ed.: *Rationalism, Empiricism and Idealism*. New York: Oxford, 1986. 179 p. Cloth $37, paper $13.50.

KNUUTTILA, SIMO and JAAKKO HINTIKKA, eds.: *The Logic of Being*. Boston: D. Reidel, 1986. xvi, 300 p. $54.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

KORNBLITH, HILARY, ed.: *Naturalizing Epistemology*. Cambridge, Mass.: MIT Press, 1985. vii, 303 p. $22.50.

LAYCOCK, STEVEN W., and JAMES G. HART, eds.: *Essays in Phenomenological Theology*. Albany: SUNY Press, 1986. 217 p. Cloth $39.50, paper $12.95.

LEINFELLNER, WERNER, and FRANZ M. WUKETITS, eds.: *The Tasks of Contemporary Philosophy*. Boston: D. Reidel, 1986. 565 p. $89.

Le PORE, ERNEST, and BRIAN McLAUGHLIN, eds.: *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. New York: Basil Blackwell, 1986. xiii, 507 p. $45.

Le PORE, ERNEST, and BRIAN McLAUGHLIN, eds.: *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. New York: Basil Blackwell, 1986. xiii, 507 p. $45.

LUCAS, GEORGE R., JR, ed.: *Hegel and Whitehead: Contemporary Perspectives on Systematic Philosophy*. Albany: SUNY Press, 1986. vi, 325 p. Cloth $44.50, paper $18.95.

LUCAS, HANS-CHRISTIAN, and OTTO PÖGGELER, eds.: *Hegels Rechtsphilosophie im Zusammenhang der europäischen Verfassungsgeschichte*. Stuttgart: Frommann-Holzboog, 1986. 544 p. 168 DM.

MACKINNON, BARBARA: *American Philosophy: A Historical Anthology*. Albany: SUNY Press, 1985. xv, 688 p. Cloth $49.50, paper $16.95.

MacLEAN, DOUGLAS, ed.: *Values at Risk*. Totowa, N.J.: Rowman & Littlefield, 1986. ix, 178 p. Cloth $28.50, paper $15.95.

MARGOLIS, J., M. KRAUSZ, and R. M. BURIAN, eds.: *Rationality, Relativism and the Human Sciences*. Boston: Nijhoff, 1986. viii, 243 p. $58.90.

MARKLEY, ROBERT, RONALD SCHLEIFER, eds.: *Kierkegaard and Literature: Irony, Repetition, and Criticism*. Norman: Oklahoma UP, 1984. xvi, 224 p. $22.95.

MARSELLA, AHTHONY J., GEORGE DEVOS, and FRANCIS L. K. HSU, eds.: *Culture and Self: Asian and Western Perceptions*. New York: Tavistock/Methuen, 1985. 321 p. Cloth $25, paper $12.95.

MARTIN, MIKE W., ed.: *Self-Deception and Self-Understanding*. Lawrence: Kansas UP, 1985. x, 316 p. $29.95.

MARTINICH, A. P., ed.: *The Philosophy of Language*. New York: Oxford, 1985. 492 p. $24.95.

MORREALL, JOHN, ed.: *The Philosophy of Laughter and Humor*. Albany: SUNY Press, 1986. xi, 270 p. Cloth $34.50, paper $10.95.

MUELLER-VOLLMER, KURT, ed.: *The Hermeneutics Reader*. New York: Continuum, 1985. xi, 380 p. $27.50.

MUSCHAMP, DAVID, ed.: *Political Thinkers*. New York: St. Martin's, 1986. ix, 259 p. $29.95.

NAGL, LUDWIG, and RICHARD HEINRICH, eds.: *Wo steht die analytische Philosophie heute?* Vienna: Oldenbourg, 1986. 192 p.

NELSON, JOHN S., ed.: *Tradition, Interpretation, and Science: Political Theory in the American Academy*. Albany: SUNY Press, 1986. xvii, 372 p. Cloth $44.50, paper $16.95.

NORTON, BRYAN G., ed.: *The Preservation of Species*. Princeton, N.J.: University Press, 1986. xi, 305 p. $29.50.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIV, NO. 5, MAY 1987

KANTIAN ETHICS AND SUPEREROGATION*

> Don't think me arrogant for saying this, but the commandments of morality are too trifling for me; for I should gladly do twice as much as they command. . .I console myself often with the thought that since the practice of morality is so bound up with sensuality, it can only count for this world, and with that thought I could still hope not to have to live another life of empty vegetating and of so few and easy moral demands after this life.[1]
>
> —Maria von Herbert

ALTHOUGH von Herbert was, we hope, unusual in finding this alleged problem to be a source of intense personal despair, many others have shared the sense, however vague, that Kant's ethics is in trouble here and that the trouble is traceable to the theoretical framework. The problem is diagnosed in two different ways. Some, like von Herbert, believe that Kant's ethics asks too little, requiring only minimal decency and saying nothing about the "higher flights of morality."[2] Others, reading the texts differ-

* I am indebted to too many people for their helpful comments on earlier drafts of this paper for it to be possible to cite all of them here. Special thanks are due to Richard Henson, who commented on a shorter version at the 1985 Eastern Division meetings of the American Philosophical Association; Laura Melim and Craig Ihara, who commented on another version at the Conference on the Virtues at the University of San Diego (February 1986); and Karl Ameriks, Alan Donagan, and Peter Simonson for extensive written comments. The paper benefited from colloquium discussions at San Francisco State University (April 1985), Stanford University (June 1985), the University of Chicago (October 1986), and the University of Illinois at Urbana-Champaign (October 1986), as well as at the APA meetings and the Conference on the Virtues. Finally, I would like to express my gratitude to the American Council of Learned Societies for its financial support in 1984/85, when I began writing this paper.

[1] From *Kant: Philosophical Correspondence 1759–99*, Arnulf Zweig, ed. (Chicago: University Press, 1967), pp. 201/2.

[2] The phrase is from J. O. Urmson, "Saints and Heroes," in A. I. Melden, ed., *Essays in Moral Philosophy* (Seattle: University of Washington Press, 1958).

0022-362X/87/8405/0237$02.60 © 1987 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ently, believe that his theory asks too much, demanding total devotion to morality and treating everything worth doing (and perhaps more) as a duty. But, despite their differences, the two sets of critics concur in taking the central problem to be Kant's failure to recognize the category of the supererogatory. Duty, it is said, is just a small part of morality (if indeed it is any part of morality). An ethical theory in which duty takes over is a severely impoverished theory.[3]

The claim is best understood in light of a set of criticisms to the effect that Kant's ethics and, more generally, Kantian ethics leave little room for friendship, fellow-feeling, loyalty, a sense of community, personal attachment and, indeed, personality. Such objections have been put forth by Julia Annas, Lawrence Blum, Alasdair MacIntyre, Michael Slote, Michael Stocker, Bernard Williams, Susan Wolf, and others.[4] A number of critics think that the underlying problem is the enormous role that the notion of duty plays in Kant's ethics. His ethics is too minimal because "duty" looms too large. There are in fact two related criticisms here. First, too much value (or too exclusive a value) is assigned to the motive of duty. Second, too much is regarded as part of our duty. The scope of duty is too broad; it should be narrowed so that part of moral conduct can then be seen as beyond duty, i.e., as supererogatory.

It is the latter objection that my paper addresses.[5] While others

[3] See Lawrence Blum, *Friendship, Altruism and Morality* (Boston: Routledge & Kegan Paul, 1980); Paul Eisenberg, "Basic Ethical Categories of Kant's *Tugendlehre*," *American Philosophical Quarterly*, III, 4 (October 1966): 255–269; Roderick Chisholm, "Supererogation and Offense: A Conceptual Scheme for Ethics," *Ratio*, V, 1 (June 1963): 1–14; Joel Feinberg, "Supererogation and Rules," in his *Doing and Deserving* (Princeton, N.J.: University Press, 1970); Philippa Foot, *Virtues and Vices* (New York: Basil Blackwell, 1978); David Heyd, *Supererogation: Its Status in Ethical Theory* (New York: Cambridge, 1982); Michael Slote, *Goods and Virtues* (New York: Oxford, 1983); Michael Stocker, "The Schizophrenia of Modern Ethical Theories," this JOURNAL, LXIII, 4 (Aug. 12, 1976): 453–466; Urmson, "Saints and Heroes," *op. cit.;* Bernard Williams, *Moral Luck* (New York: Cambridge, 1981); and Susan Wolf, "The Superficiality of Duty," unpublished manuscript. This list is far from exhaustive.

[4] See, in addition to the works by Blum, Slote, Stocker, Williams, and Wolf listed above, Alasdair MacIntyre, *After Virtue* (Notre Dame, Ind.: University Press, 1981), and "Is Patriotism a Virtue?" (Lawrence: University of Kansas Philosophy Department, 1984); and Julia Annas, "Personal Love and Kantian Ethics in *Effi Briest*," *Philosophy and Literature* (April 1984): 15–31. I discuss Annas's paper in my "Was Effi Briest a Victim of Kantian Morality?", forthcoming in *Philosophy and Literature*.

[5] For a discussion of the former objection, see Barbara Herman, "On the Value of Acting from the Motive of Duty," *Philosophical Review*, LXXXII, 3 (July 1981): 359–382. My "The Alleged Moral Repugnance of Acting from Duty" [this JOURNAL, LXXXI, 4 (April 1984): 197–220] also discusses that objection, but without seeking to defend Kant's particular conception of acting from duty. In sec. II of that essay I briefly discuss the other criticism: that Kantian ethics is too restricted to morality's requirements; I now think that discussion of it is inadequate.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

have sought to defend Kant by showing that in fact his theory does leave room for the category of the supererogatory, I argue that it does not, but that this does not constitute a flaw in his theory. Unlike Kant's other defenders, I argue against the thesis that any ethical theory that does not recognize a special category of the supererogatory is *ipso facto* flawed,[6] and suggest that Kant's treatment of the category of acts that contemporary ethicists speak of as supererogatory may well be preferable to the supererogationist approach.[7]

The first section of my paper attempts to explain what supererogationists mean by 'supererogation' and by the claim that any ethical theory that fails to accommodate it is faulty. The second section asks whether Kant's ethics leaves room for the supererogatory. Finding that it cannot, I then investigate Kant's apparent reasons for not accommodating it and suggest that the considerations he adduces should give us pause, and prompt us to question the supererogationist thesis. Finally, I ask why it is supposed to be so important to have a special category of the supererogatory and argue that the reasons that supererogationists put forth offer no more support for their claim than for following Kant's lead and distinguishing between perfect and imperfect duties.

I. WHAT SUPEREROGATIONISTS BELIEVE

Supererogationists believe that there are acts that fit the following criteria: they are beyond duty and they are morally good and praiseworthy. They are beyond duty in that, as David Heyd puts it, "they fulfil more than is required, over and above what the agent is required or expected to do," and thus are optional (*op. cit.*, 1).

Stated thus, the thesis may seem to be about *what exists, viz.* acts of a certain sort. Sometimes supererogationists use terminology that supports this reading of the thesis. Heyd, for example, writes that "supererogationism is justified by showing that some supererogatory acts must exist" (166). But it is clear from the rest of the sentence: "because society cannot require of the individual every act that would promote the general good," that he is not thinking of the thesis as a claim about what exists or, less clumsily, as a claim about what happens or what people do. In any event, what happens or what people do is clearly not the issue. No one is questioning whether people do ever donate blood in a voluntary system, give money to the

[6] I do not claim that *no* ethical theory needs the category. It may be important for utilitarianism and other ethical theories that emphasize the maximization of some good. Whether or not it is will not be considered here.

[7] I apologize for using that abomination of a word, 'supererogationist'. I follow Heyd in using it, since it is a convenient term, and I think it preferable to 'supererogatorian' (listed as obsolete in the *Oxford English Dictionary*).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

poor, or forgive a person for the wrong he has done (to take some of Heyd's examples of supererogatory acts) (2).

The issue is not about what people do, but about how to view such acts as those just listed, i.e., about how to classify them. Thus Heyd explains, in the introduction to his *Supererogation: Its Status in Ethical Theory,* that one of the book's aims is "to show that supererogation is a distinct and determinate class of moral action deserving of moral scrutiny, and that there are both theoretical and moral reasons for treating these acts as belonging to a separate category" (1). He emphasizes that "one criterion of the acceptability of a theory" is its "treatment of supererogation" (10), and says that supererogation cannot "be accounted for. . .by a theory which takes duty as exhausting the whole realm of moral behaviour" (3). Similarly, Michael Clark writes that "the notion of the supererogatory is not definable solely in terms of what we ought to do."[8] The thought here is that ethical theories often fail to do justice to the category of the supererogatory, and Kant's, supererogationists claim, is one such failure. For somewhat different reasons, J. O. Urmson takes the same position. In his classic "Saints and Heroes," he calls for a two-tier ethical theory that sharply separates the supererogatory from duty: "Thus as moral theorists we need to discover some theory that will allow for both absolute duties, which, in Mill's phrase, can be exacted from a man like a debt, to omit which is to do wrong and to deserve censure, and that may be embodied in formal rules or principles, and also for a range of actions which are of moral value and which an agent may feel called upon to perform, but which cannot be demanded and whose omission cannot be called wrongdoing" (208).[9]

In short, the supererogationist holds it to be a condition of adequacy for any ethical theory that it not only recognize the possibility (and actuality) of certain acts that go beyond what is morally required and are morally praiseworthy, but also that it treat these acts as constituting an independent category, independent in the sense that their value cannot be explained in terms of duty. The claim is about how to view such acts, and how to categorize them within ethical theory.

On what grounds can one argue that a category is or is not needed in ethical theory? The territory is murky and ill-charted, but I think

[8] "The Meritorious and the Mandatory," *Proceedings of the Aristotelian Society,* new series, LXXIX (1978/79), p. 24.

[9] I say that Urmson takes the same position "for somewhat different reasons" because, unlike Heyd, he (apparently) does not realize that the acts he thinks of as supererogatory can be classified in Kant's ethical theory other than as merely permissible or as severally required. (See below, sec. II.) Heyd sees this, but is still dissatisfied with Kant's treatment.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that we can say this much: a category is needed in ethical theory either for its explanatory value, or because it is practically valuable, valuable for (among other things) self-direction, moral education, or the flourishing of a community.[10]

In denying the supererogationist thesis, I shall first of all be asserting that the value of supererogatory acts *can* be explained in terms of duty, and *is* thus explained by Kant. Supererogationists will no doubt respond that I do not recognize their special value, and they are right (assuming that 'recognize' is not here taken as a success verb). In addition, I will be claiming that all the (meaningful) theoretical work accomplished by the category of the "supererogatory" can be handled by the Kantian framework, and perhaps handled better.[11] I will not be denying that there are acts that can (contextually, and as acts; not as act-types) be labeled "supererogatory." But it is the wrong thing to focus on; it is not something that ethical theory should emphasize. Supererogatory acts "exist," but they do not form an ethically useful or theoretically interesting kind.

With these preliminaries out of the way, I turn to the question of whether Kant's ethics leaves room for the supererogatory. I begin by examining an argument put forth by Thomas E. Hill, Jr., from which I will build my case that Kant had no need for a special category of supererogatory acts.[12]

II. HILL'S ARGUMENT

In a very illuminating article, "Kant on Imperfect Duty and Supererogation," Hill argues that an examination of Kant's classification of duties reveals that he can recognize the category of supererogatory acts.[13] Although I like his reading (and tidying up) of Kant's classification and agree that it shows Kant's ethics to be, as Hill says, "less rigoristic than commonly thought" (62), I do not think that he

[10] Indeed, these are the sorts of considerations Heyd offers in support of supererogationism. See sec. IV, below.

[11] For an indication of why I add 'meaningful' see fn 48, below.

[12] Another Kant scholar who seeks to defend Kant by arguing that he does have room for the category of the supererogatory is Onora O'Neill (formerly Nell). She writes in a short section of her *Acting on Principle: An Essay on Kantian Ethics* (New York: Columbia, 1975), "If one defines supererogatory acts as those which go beyond duty in the sense of doing *more* than is obligatory, then it would seem that acts which are obligatory and also morally worthy are acts of supererogation. If, on the other hand, like Eisenberg, one defines supererogatory acts as acts which though not obligatory are morally worthy, then Kant also can allow for these" (pp. 94–96). But this misses the mark. For Kant, an act is morally worthy if and only if it is done from duty, whereas supererogatory acts are supposed to be especially good not thanks to the agent's motive in performing them, but because they are strictly optional (and yet morally significant).

[13] "Kant on Imperfect Duty and Supererogation," *Kant-Studien*, LXXII (1971): 55–76.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

establishes that it can accommodate the supererogatory. In the end, he can locate a space for the supererogatory only by pressing Kant's ethics into a mold alien to Kant, though very familiar to twentieth-century philosophers. In itself this is not an objection, since it is certainly valuable to propose a friendly amendment to a major thinker's view, i.e., an alteration that improves the view while leaving most of it intact. I will argue, however, that it is by no means clear either that the proposed modification improves Kant's ethics or that it leaves most of it intact.

Hill locates the class of supererogatory acts among a subset of Kant's imperfect duties which Hill designates the 'wide imperfect duties'.[14] Wide imperfect duties (which, as Hill suggests, are the paradigmatic imperfect duties) differ from perfect duties in two respects. First, wide imperfect duties—indeed, all imperfect duties—are primarily duties to adopt a maxim (or what comes to the same thing, to embrace an end). As Hill says, "One could fulfill a perfect duty by doing a right sort of act without having adopted any principle relevant to the case but one could not fulfill an imperfect duty in this manner" (62). So, for example, if I have a perfect duty not to make false promises, I fulfill that duty even if I refrain only because I think it a prudent policy never to make false promises. But I do not fulfill my imperfect duty of promoting the happiness of others if I do good deeds for others only with the aim of impressing them or winning their favor.

Second, wide imperfect duties allow latitude not allowed to perfect duties. They permit latitude not only in that they set very broad parameters for satisfying a principle. They also allow us, in Hill's words, "freedom to choose to do x or not on a given occasion, as one pleases, even though one knows that x is the sort of act that falls under the principle, provided that one is ready to perform acts of that sort on some other occasions" (61).[15] Thus, a principle of wide imperfect duty, e.g., a principle of beneficence, not only leaves open what it is that I am to do when I act beneficently; it also permits me to omit altogether to act beneficently in certain circumstances in which I have an opportunity to help someone, provided that I act benefi-

[14] Kant's terminology is admittedly looser than this, as Hill recognizes. The exegetical difficulties are brought out in Mary Gregor, *Laws of Freedom* (New York: Basil Blackwell, 1963), ch. 7, as well as in Hill's paper.

[15] I have simplified his account. Hill actually distinguishes three types of latitude. One type allows "room for judgment in deciding whether or not a given principle is relevant to a particular situation"; another permits "freedom to choose various ways of satisfying a principle in a particular situation once we decide that the principle applies"; and the third is as quoted in the text. Wide imperfect duties allow all three types of latitude; narrow imperfect duties allow only the first two types.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

cently in other circumstances, and that I really *have* adopted a maxim of beneficence. Similarly for a principle of developing my talents. In contrast, principles of "narrow" imperfect duties, e.g., duties to respect others, allow only the first type of latitude.

To sum up: imperfect duties, unlike perfect duties, are primarily duties to adopt a maxim (or an end). Indirectly they require us to act accordingly, but exactly what we are to do cannot be spelled out (not because Kant leaves it unclear, but because it is in the nature of imperfect duties that this cannot be spelled out). Principles of wide imperfect duties allow even greater latitude (or, more precisely, latitude of a different sort) than do those of narrow imperfect duties. They allow us to omit some actions that fall under the principle in question.

It should be apparent that wide imperfect duties bear a promising resemblance to supererogatory acts. Still, Hill suggests, for an act to be supererogatory it is not enough that it fulfill a principle of wide imperfect duty, since there are circumstances where such an act will be mandatory ("e.g., if one who has continually neglected to help others is faced with his last opportunity").[16] The best candidate for a supererogatory act, Hill says, is an act that

(a) is of a sort commended by a principle of wider imperfect duty,
(b) is motivated by a sense of duty (or, perhaps, respect for moral reasons),
(c) is neither forbidden nor required by another, more stringent duty. . .,
(d) is in a context where no alternative is required by more stringent duty and there is at least one alternative that is neither forbidden by more stringent duty nor commanded by other principles of wide duty, and
(e) is done by an agent who has adopted the relevant principle of wider imperfect duty and has often and continually acted on that principle (71).

Has Hill shown that Kant has a place for supererogatory acts? It depends on what is meant by saying that philosopher *X* has a place for supererogatory acts. The speaker might mean that philosopher *X* can, without inconsistency, recognize that there are acts that, in

[16] There is a puzzle here: If I have continually neglected to help others, can I fulfill a principle of wide imperfect duty by helping someone now, when I see that it is my last chance? If I have never helped anyone before, it is hard to claim that I have adopted a maxim of beneficence. (It is not as if one could easily just fail to come across, or hear of, someone in need. One would have to go to considerable trouble to avoid occasions for acting beneficently.) I would have to have undergone a moral rebirth for my act of helping others, in the example, to fulfill an imperfect duty.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

certain circumstances, it is good but not morally required to perform. Alternatively, the speaker may be claiming something more, namely that philosopher *X* can, without inconsistency, treat the category of the supererogatory as supererogationists ask that it be treated. In the first and weaker sense, Hill has shown that Kant has a place for supererogatory acts. Even this needs qualification: Hill has shown that there are acts which it would be good to do but which one is not morally required to do, provided that 'acts' is understood to mean particular acts, not act-types. To the extent that Kant can recognize supererogatory acts, these are acts which can only be specified by reference to the particular context, including not only the agent's motives but also her past performance and her principles. We cannot, on the Kantian scheme as explicated by Hill, say that acts of helping others or acts of heroism are as such supererogatory; whether or not they are depends on the factors just mentioned.

Before taking measure of the gap between what Hill has shown and what supererogationists want, we should note that establishing even the fairly weak claim that Hill has established is no minor feat. It overturns Urmson's claim that "this threefold classification"—of the morally required, the permitted, and the forbidden—"or any classification that is merely a variation on or elaboration of it, is totally inadequate to the facts of morality," a claim which he defends by drawing attention to saintly and heroic acts (198/9). Hill's analysis shows that Urmson is just wrong if he supposes that such acts cannot fit anywhere in Kant's classification. They count, in Kant's scheme, as acts which, while not duties, are ways of fulfilling imperfect duties. Hill's discussion also shows Kant's classification to be in at least one important respect richer than Urmson's. In addition to providing a place for the type of acts just mentioned, Kant's classification allows us to distinguish between duties which can be exacted from one (the juridical duties—and the only type of duty Urmson recognizes) and those which cannot be (the ethical duties).

All this notwithstanding, to show that Kant can recognize supererogatory acts in this weak sense is not yet to satisfy the supererogationists. Most would object to Hill's conditions (b) and (e) on the grounds that they overemphasize acting on principle and from duty. We might try revising Hill's criteria accordingly. But in addition to being an unpromising maneuver [since (e) would have to be altered in a way that would not square with the plan to locate the supererogatory among imperfect duties], it would at best solve only a small part of the problem. The real difficulty is that supererogationists are trying to get away from duty, but the Kantian analysis keeps pulling what is supposed to be "beyond" duty back under the rubric of duty.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Kant treats the acts in question not as beyond duty, but as ways of fulfilling one's imperfect duties. To satisfy the supererogationist, one would have to show that Kant does not just have a cubbyhole somewhere for supererogatory acts, but is able to accord them a proper place, and thus proper regard.

Hill realizes this, I think, for he notes that condition (b) seems to require that "one can do something supererogatory only if he mistakenly thinks that it is his duty." He proposes a modification of Kant's position: "In reply we might suggest that it is not an unreasonable extension of Kant's position to say that what is required for moral worth is not a motive to do one's *duty* but a motive to do what is demanded or encouraged by moral considerations." The nature of the proposal becomes clearer when he takes note of a similar difficulty concerning (e). "How, one may wonder, can I be going 'beyond duty' if I am guided by a principle of duty?" His response underscores his view that Kant's terminology does not aptly express his true position.

> Kant's terminology of "imperfect" and "perfect" duty does confuse the issues. It is as if Kant started to work out a moral theory on the model of legal-like strict duties, and then, discovering that there is more to morality than duty, still retained the old labels for types of duty rather than spoil the symmetry of his theory by changing to more natural expressions. For example, what Kant is concerned to say about beneficence is (i) it is a duty to adopt a maxim of beneficence, and therefore (ii) it is a duty to promote the happiness of others sometimes, but also (iii) when one has satisfied these minimum and rather indefinite requirements, one may promote their happiness or not, as one pleases, but to do so with the proper motive will always be of positive moral worth. Kant tried to say all of this with his restrictive terminology of duty when it could be put more simply by making an early distinction between what is obligatory and what is merely good to do (74).

Hill is suggesting that Kant would do better to distinguish between what is obligatory and what is merely good to do. His proposal would have the effect of narrowing the scope of duty so that certain acts which Kant sees as fulfilling our imperfect duties count instead as supererogatory. The proposal might still not satisfy supererogationists. Since (e) is left intact and (b) is loosened only slightly, the supererogatory act may seem to be robbed, by this Kantian transformation, of its special goodness. Rather than try to resolve this matter, let us focus on the question of whether Kant would (or should) welcome Hill's proposed revision, a revision which would at least go *some* way toward satisfying the supererogationist.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

III. WHY KANT WOULD REJECT HILL'S PROPOSAL

One set of reasons for thinking that Kant would reject Hill's proposal can be gleaned from the Methodology of the *Critique of Practical Reason.*[17] Kant is concerned in that section with the question of how "we can make the objectively practical reason also subjectively practical" or, in other words, how "we can secure to the laws of pure practical reason access to the human mind and an influence on its maxims" (155 [151]). After proposing that "educators of youth" present their subjects with examples from biographies, so that "by comparing similar actions under various circumstances, they could begin to exercise the moral judgments of their pupils in marking the greater or less moral significance of the actions," Kant remarks:

> But I wish they would spare them examples of so-called noble (supermeritorious) actions, which so fill our sentimental writings, and would refer everything to duty only and the worth which a man can and must give himself in his own eyes through the consciousness of not having transgressed his duty, since whatever runs up into empty wishes and longings for unattainable perfection produces mere heroes of romance, who, while priding themselves on their feeling of transcendent greatness, release themselves from observing the common and everyday responsibility as petty and insignificant [K2 158/59 (155)].

Two concerns of relevance to our discussion are evident in the passage just cited. First, Kant worries, if emphasis is placed on heroic deeds, morality will seem too remote and, for the most part, optional. This is closely tied to Kant's conviction, reiterated throughout the second *Critique,* that the motive of duty is weakened by romantic enthusiasm for "actions called noble, magnanimous and meritorious" (161 [157]).[18] If identified too closely with heroic feats, morality might come to be thought of as a sort of spectator sport, or perhaps as a profession. It would be one of a number of fields in which those with a special gift for it might excel, but which is beyond the reach of most.[19] This is, of course, very much at odds with Kant's

[17] Lewis White Beck, trans. (Indianapolis: Bobbs-Merrill, 1956). Hereafter cited as "K2." Bracketed page numbers refer to the Prussian Academy edition.

[18] See also p. 87 [84/5]: "The mind is disposed to nothing but blatant moral fanaticism and exaggerated self-conceit by exhortation to actions as noble, sublime, and magnanimous. By it people are led to the illusion that the determining ground of their actions is not duty, i.e. respect for the law. . . . This law always humbles them when they follow (obey) it, but by this kind of exhortation they come to think that those actions are expected of them not because of duty but only because of their own bare merit." Cf. *Religion within the Limits of Reason Alone,* Theodore M. Greene and Hoyt H. Hudson, trans. (New York: Harper & Row, 1960), p. 44 [48/9].

[19] Interestingly, David Heyd sees it as an argument in *support* of supererogation that "supererogatory action breaks out of the impersonal and egalitarian framework of the morality of duty" (175).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

conception of morality as being within everyone's reach and incumbent on all.

The second concern is that, insofar as one manages to maintain some appreciation of that part of morality which would still fall under the heading of *duty,* one will feel that supererogatory acts substitute for fulfilling morality's requirements. Worse yet, there is a real temptation so to substitute (if one chooses supererogatory acts of the nonheroic, nonsaintly variety). One can puff up with self-satisfaction at having done something "extra" for someone; it is more difficult to feel smug and superior about doing what, one believes, anyone in those circumstances is morally required to do.

It is tempting to dismiss Kant's worries as farfetched, and more worthy of the attention of biographers than that of philosophers interested in evaluating the category of supererogatory acts. This impression fades somewhat as we notice the extent to which our own society confirms Kant's fears.[20] According to popular, although usually unarticulated dogma, I need not feel guilty about treating my wife more like a pet than a person since I buy her expensive clothes; and (something of an extension of this) we need not worry too much about how badly underpaid secretaries are since we have "Secretaries' Week"! (And it is much more fun to give those in an inferior role a present than to try to bring it about that they cease to occupy an inferior position.) To take a different example, we all know that extreme poverty and social injustice abound, and most of us believe that they are not ineradicable; but it is much easier, and at the same time gratifying, to offer some surplus cheese to the poor than to work to alter the conditions that make poverty and inequality social institutions.[21]

[20] The notion that the special merit of supererogatory acts could cancel out the moral "debt" incurred by a failure to fulfill a moral requirement was implicit in medieval Church doctrine, according to which the penitent could be absolved without fulfilling his penance (provided that he gave money to the Church, joined the Crusades, or did whatever else the current conditions for indulgences stipulated). Aquinas and Bonaventure emphasized that the penance was not canceled but paid: the debt was paid from the Church's treasury, which stored the extra merit from the supererogatory acts of Jesus, Mary, and the saints. See P. F. Palmer, "Indulgences," in the *New Catholic Encyclopedia,* vol. VII (1967), pp. 482–484. Heyd's *Supererogation* contains a fascinating chapter on the theological issues concerning supererogation. As he points out, Kant's thinking was undoubtedly influenced by Martin Luther's attack not only on the system of indulgences, but on the underlying doctrine of supererogation.

[21] Cf. Mordecai Nisan, "Moral Balance—A Model for Moral Decisions," in W. Edelstein and G. Nunner-Winkler, eds., *Zur Bestimmung der moralphilosophische und socialwissenschaffliche Beiträge zur Moralforschung* (Frankfurt: Surkamp, 1986). Nisan argues that our moral deliberations reflect a point-system conception of morality.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

In developing this point I have in mind a passage from the *Doctrine of Virtue*. In the section on beneficence Kant queries:

> The ability to practice beneficence, which depends on property, follows largely from the injustice of the government, which favours certain men and so introduces an inequality of wealth that makes others need help. This being the case, does the rich man's help to the needy, on which he so readily prides himself as something meritorious, really deserve to be called beneficence at all?[22]

I take it that Kant would be all the more worried if alongside of (or instead of) the imperfect duties we had a category of supererogatory acts. For although the imperfect/perfect distinction may make it easy for us to regard some of our duties as wide and imperfect when, in fact, the sociohistorical underpinnings of the situation indicate that we may well have something closer to a duty of justice (a perfect duty), at least it offers no special temptation to dismiss them as not really duties at all.

It may seem unfair to supererogationists to allude to ubiquitous abuses of the notion, even if they are the very sorts of abuse Kant worried about. After all, an argument against an abuse of a notion is not a good argument in favor of giving up the notion entirely. If we firmly reject the suggestion that one stores up "credit" through one's supererogatory acts and if heroism is not emphasized too much in moral education, why suppose that we need to abandon the category of the supererogatory?

Although this argument has merit, it should also be borne in mind that it is not an easy matter to reject—as a culture—a self-serving belief that we never even articulate to ourselves; in this case, the notion that good deeds make up for failure to take morality's requirements seriously.[23] And in a society like ours which encourages self-indulgence, self-flattery, and competitiveness, a "point system"

[22] Kant, *The Doctrine of Virtue*, Mary J. Gregor, trans. (Philadelphia: Pennsylvania UP, 1964), p. 122 [454]. Hereafter cited as "DV." See also fn 1 in the Methodology (K2), Kant's *Education* (Ann Arbor: University of Michigan Press, 1960), pp. 104/5 [490–491], and "Duties towards Others" in his *Lectures on Ethics*.

[23] Nor is it clear that supererogationists always want this to be rejected. Indeed, the only contemporary book devoted to the subject—and a very well argued and scholarly work—suggests that, although we should reject one platform of the Church doctrine on supererogation, i.e., that moral "credit" can be transferred from one person to another, "the model itself may serve—*mutatis mutandis*—as an illustration of what is meant by 'merit', namely, a certain surplus which is credited to the agent." See Heyd, p. 140. Thus, the central idea that the agent has earned "extra credit" persists.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of morality, with its suggestion that one can (especially if one has money) buy one's way out of morality's requirements through acts of "generosity," may be difficult to resist.[24] Thus, although they do not show that we must reject the category of the supererogatory, these considerations provide reason for insisting on hearing just why it is supposed to be so important to recognize a special category of the supererogatory. But, before canvassing the supererogationist arguments, we have other Kantian objections to consider.

The second set of objections brings us to the heart of the matter, for they oppose dividing what Kant means by duty into what one really must do and what, as Hill puts its, it would be good to do. Kant's opposition to such fragmentation is not, to my knowledge, ever explicitly stated, but it can easily be extrapolated from the texts. First, and related to the objections mentioned above, such fragmentation makes much of morality *optional*. It invites what might be called "yuppie ethics," for it introduces a division, a fence around what "I may choose, as I please," separating it from what I "have to" do.[25] I have done "my duty" (conceived, following Hill's suggestion, more narrowly than Kant wanted it); now my time, my choices are all mine (except, of course, for the constraints of perfect duties). Any other attention to morality is, for me, strictly optional.

This is all quite alien to Kant's ethics. There is no clear line of demarcation between what I must do, morally, and what is nice but morally optional. Nor does Kant attempt to trace such a line of demarcation. To do so he would have to give up a very central, even foundational, duty: the imperfect duty we each have to strive to perfect ourselves morally. This duty, which we cannot escape by being "good enough," underlies our other imperfect duties. To give it up would be to alter, quite radically, the structure of Kant's moral scheme.[26]

Now this may prompt some to say, "So much the worse for Kant. It does indeed look as if he cannot recognize a realm of acts that are good to do but beyond duty; but that only confirms my belief that I should not bother with Kant. It does not shake my belief in the importance of the supererogatory." I want to suggest that the Kantian picture, with its rejection of the distinction between what I ought to do and what is good to do but strictly optional, merits serious

[24] Cf. Feinberg's discussion of a point-system conception of morality in "Supererogation and Rules." See also Nisan, *op. cit.*

[25] I take no credit for this label. I heard it from Onora O'Neill, who thinks that it originated with Dale Jamieson.

[26] DV, p. 113 [445]. I am indebted to Christine Korsgaard for calling my attention to this problem.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

consideration. As Warner Wick remarks,[27] the imperfect duties "not only allow us considerable leeway in achieving their objectives, but they are indefinite in the further sense that what they demand is also without assignable limits. We can never say, 'There, I have at last done all that I ought to do for other people'." I regard this as an advantage of the Kantian view, though I realize that to some it would seem a disadvantage.

The central difference between these two approaches lies not in the deeds required, but in the fact that imperfect duties are primarily duties to have certain maxims (the having of which will then direct one's actions to some extent, but in no very determinate way). The difference can be brought out by noting a subtle tension in Hill's account of what Kant "is concerned to say about beneficence." After saying that "it is a duty," on Kant's view, "to adopt a maxim of beneficence," and that this entails a duty of promoting the happiness of others sometimes, Hill says that, according to Kant, "when one has satisfied these minimum and rather indefinite requirements, one may promote their happiness or not, as one pleases" (74). But it is not entirely as one pleases; if it were, virtue would be optional, and the duty to improve oneself would be incumbent only on those who, morally speaking, are especially derelict. Morality would indeed be minimal.

On Kant's picture one cannot both have a maxim of beneficence and regard one's "work" as over. This is not because having the maxim of beneficence entails the odd mind-set of Susan Wolf's moral saints, who strive to do as much as possible and to make their every action as virtuous as possible.[28] But, although a maxim of beneficence does not require that one seek to maximize with respect to the happiness of others, neither does it permit one to say, "Ah, I have met my weekly quota; now I can relax." Having this maxim is incompatible with drawing a line around one's "own life" or "own projects" and seeing morality the way one might see mowing the lawn (or as a child might see doing his homework or performing his boyscout deeds): as something to get out of the way.

In reply to the above objection, one might suggest that we can avoid the egocentric, yuppie approach without completely rebuffing the supererogationist. We can do so if we retain Kant's category of imperfect duties for routine, nonheroic beneficent acts, while relegating to the supererogatory the spectacular deeds (on which Urm-

[27] See his introduction in Kant, *Ethical Philosophy* (Indianapolis: Hackett, 1983), p. li.

[28] "Moral Saints," this JOURNAL, LXXIX, 8 (August 1982): 419–439. See also R. M. Adams's reply, "Saints," this JOURNAL, LXXXI, 7 (July 1984): 392–401.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

son laid so much emphasis): hurling oneself onto a live hand grenade in order to protect one's buddies, rushing into a burning building in the hopes of saving the people in it, leaving one's home and comfortable medical practice to work in a plague-ridden area, etc.[29] This would, in effect, narrow the scope of the supererogatory, but at least recognize the category.

This revised proposal is, I think, an improvement over the original. Yet, although it preserves the notion that we continue to have a duty to strive to be virtuous even if we are, morally speaking, well above average, it limits that duty in a different way. For it rests on an assumption that, if acts are only hard enough for us, they should be regarded as optional.[30] Many people no doubt accept this assumption and believe that any decent ethical theory will honor it. (By a contortion of the overused and misunderstood principle, "Ought implies can," the assumption may even seem unassailable.) But there are good reasons, which Elizabeth Pybus develops in a discussion of Urmson's "Saints and Heroes,"[31] for questioning the wisdom of building such an assumption into moral theory. Pybus notes that there are a great many acts (not just those for which people are given special medals) which require considerable courage. Take, for instance, the agoraphobe who does an elderly neighbor's shopping. She "is just as brave as the heroic soldier" and "might even find the death threatened by the grenade less frightening." In mulling over the issue of whether we should recognize in ethical theory a special category of the supererogatory (and whether a community should recognize it), it is important, Pybus emphasizes, to get the question right.

> We should not ask whether the action of throwing oneself on a grenade is beyond the call of duty, but whether actions of a certain sort, viz., very brave ones, are beyond the call of duty. And they are not. Clearly we cannot slide out of doing our duty by saying that we are not brave enough. Sometimes we may be excused for a loss of nerve, but we cannot remain cowards all our lives, and use that as an acceptable excuse for fulfilling only the basic requirements of morality (198).

The case is all the clearer when we look at other moral requirements: e.g., to be just or beneficent. Imagine someone saying, "I am sorry; I am just not the giving type. That may be easy for others; I am

[29] The first and third examples are from Urmson.
[30] It does not logically require the assumption, but its motivation appears to derive from it.
[31] "Saints and Heroes," *Philosophy*, LVII, 220 (1982): 193–199. In citing this passage I am uneasy about 'basic'. That I would have put the word in scare quotes may reveal a difference between Pybus's views and mine.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

just selfish, so you cannot expect me to help you in times of need." Or: "Fairness is not my forte. It is unreasonable to expect it of me when it is so hard for me, so contrary to my nature. Give me something easier to do." As Pybus points out—and as Kant held—morality is pervasive: we cannot just complete certain chores (picking the ones that best fit our temperament), and then be off the hook.

Although saintly and heroic acts are not severally required of us, we are not permitted to make a point of fulfilling our imperfect duties in as easy a way as possible. Indeed, to permit myself to forego all acts that are very difficult for me would, on Kant's view, be to undermine my freedom (more specifically, to act as if my particular fears, desires, and aversions positively preclude that I perform certain acts, and therefore to act as if I am less free than I really am). John Procter, in Arthur Miller's *The Crucible,*[32] refuses on pain of death to provide a public "confession" to witchcraft, realizing that to do so would lend an air of legitimacy to the Salem witch trials. ("I have three children—how may I teach them to walk like men in the world, and I sold my friends?") We may not know whether we would act similarly in such circumstances, but we know that we could, and in recognizing that we are capable of so acting we recognize our freedom. Kant says the following concerning someone asked to imagine that he is threatened with death by his sovereign unless he makes "a false deposition against an honorable man whom the ruler wished to destroy under a plausible pretext": "that it would be possible for him [to refuse] he would certainly admit without hesitation. He judges, therefore, that he can do something because he knows that he ought, and he recognizes that he is free—a fact which, without the moral law, would have remained unknown to him" (30 [30]).[33]

The reason mentioned above for rejecting Hill's amendment to Kant's ethics focuses on the fact that one may not, on Kant's view, adopt a policy of never doing anything to help others which one finds extremely difficult or distasteful. In that sense, saintly and heroic acts are, as a class, not beyond duty. But there is yet another sense in which they are not beyond duty, and this consideration provides further reason for believing that Kant would reject Hill's proposal.

The sense of duty is supposed to govern all of our conduct, unifying and harmonizing our various aims and maxims. But, if we recognize as a special category acts that are beyond duty, this would tend to limit the scope, and thus the contribution, of the sense of duty.

[32] New York: Viking Press, 1953; p. 143.

[33] I am grateful to Christine Korsgaard for suggesting that I call attention to this connection between freedom and virtue.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The sense of duty alerts us to moral questions concerning admirable conduct, to issues we are otherwise likely to miss. If heroic and saintly acts are viewed as beyond duty, they seem to be, as good acts, off in a realm unto themselves. We tend not to weigh their merits against potentially competing considerations and not to ask whether the actions are in conflict with some duty. More importantly, we do not expect the agents to do this; indeed, hesitation or caution about such acts is thought of, rather, as showing a lack of bravery.[34] In the domain of nonheroic supererogatory acts there is a similar (although somewhat weaker) expectation that these acts are always good to do. After all, they are nice, well-intentioned, and (let us suppose) backed by real goodness; what could be objectionable?[35] The problem is that sometimes acts that seem and perhaps are noble and meritorious (or just plain nice) are contrary to duty. Heroic acts will sometimes violate a duty to oneself, and favors done (unbidden) for others may leave them with an unwanted sense of having burdened us, of being indebted to us, or (as when I help the child next door too much with her homework) with too little self-sufficiency.

Acts of heroism are, I think, typically thought of as off in a realm unto themselves, and it is no wonder that they are. We admire the qualities that enable the person who dives in midwinter into the Potomac in an effort to save the victims of a plane crash. And we are grateful that people with such qualities (courage, willingness to risk one's life to help others, etc.) exist. The fact that we (correctly) admire and appreciate such people is, however, no reason for thinking that their heroic actions are inevitably right. Suppose (imagining a case different from the actual one of five years ago) that the would-be rescuer is almost certain to drown, as well as not to save

[34] Traditionally, and still in vogue in some circles today, the word for this has been 'effeminacy'. The thought that it is unmanly to have second thoughts about, e.g., rushing into a burning building to rescue someone, may lurk behind the assumption that such acts are invariably good. Judging from the philosophical literature on supererogation, there appears to be a link between supererogation and the traditional male domain. The favorite examples of supererogatory acts tend to be those which exhibit a remarkable degree of some masculine virtue. Not only are feminine virtues dramatically underrepresented, Heyd (and earlier Urmson) explicitly discounts a mother's "great sacrifice for her child." It is not "strictly speaking" supererogatory, Heyd explains, because it pertains "to the sphere of natural relationships and instinctive feelings (which lie outside morality)" (Heyd, p. 134).

[35] Often (especially in ordinary moral thinking) the function of moral principles and reflection is thought to be that of a corrective to "defective" tendencies and impulses: cowardice, malice, spite, and so on. (I discuss this view in my "Morality as a Back-up System: Hume's View?", forthcoming in *Hume Studies*.) One of Kant's contributions, it seems to me, was to emphasize that moral scrutiny is needed even with respect to acts of kindness. No acts should be thought of as, in that sense, beyond duty.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

anyone. (If this is not good enough, suppose the agent is as likely to impede the person in danger of drowning as to help him.) As Kant points out, the mere fact that an act is self-sacrificial and admirable does not entail that it is good to do, that one ought ideally to do it (162 [158]).

We need, in short, to disentangle the admirable from the right. A mode of conduct may be admirable in that it points to a virtuous character trait, and yet wrong. An act may be right, though not particularly admirable. An emphasis on the supererogatory makes it difficult to keep the two distinct.

In the foregoing I have presented a number of reasons for thinking that inclusion of the category of the supererogatory in Kantian ethics would be undesirable. Although in some instances these considerations may seem pertinent only for Kantian ethics and moving perhaps only to Kant, some of the objections have, I think, broader application and appeal. Their force, however, must be weighed against considerations that support inclusion of the category of the supererogatory in ethical theory. Are there reasons for including it which outweigh the objections presented above?

IV. WHY WE ARE SUPPOSED TO NEED THE CATEGORY OF THE SUPEREROGATORY

One common argument is that an ethical theory without the category, such as Kant's, fails to accord with our common-sense intuitions and with our ordinary use of the word 'duty'. Appealing to "common opinion," Paul Eisenberg states that many of the actions or modes of conduct that Kant regards as duties "are in fact. . .acts of supererogation" (267/8).[36] One of his examples is the duty (or alleged duty) to promote the happiness of others. Appeal to common opinion is often dubious; it is decidedly suspect on issues of whether we have to promote the happiness of others or whether it is optional. This is one area where ethics should hope to do more than articulate the common opinions of what is, in our case, a rather self-centered and ungenerous society.

Other arguments turn on the importance of maintaining minimal standards of moral decency and community harmony while also recognizing personal autonomy and the need, on the part of individuals, not to be overburdened by societal and moral demands. Urmson offers the following five reasons "why our code should distinguish

[36] It has become common in the past few years to speak as if there were nothing problematic in the supposition that one condition of adequacy for moral theories is that they must not make excessive demands on us. For an insightful discussion of this notion of "excessive demands," see Norman Care, "Career Choices," *Ethics*, XCIV, 2 (January 1984): 283–302, especially sec. V.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

between basic rules, summarily set forth in simple rules and binding on all, and the higher flights of morality of which saintliness and heroism are outstanding examples":

1. It is critical that certain moral demands be given a "special status" and that "exceptional pressure" be exerted to secure compliance to them.
2. "Basic duties. . .must be. . .within the capacity of the ordinary man."
3. "A moral code. . .must be formulable. . .in rules of manageable complexity," and rules regarding when one must nurse sick neighbors or help out in leper colonies cannot be so simply formulated.
4. "A line must be drawn between what we can expect and demand from others and what we can merely hope for and receive with gratitude when we get it."
5. It is "better that pressure should not be applied" when not necessary, i.e., when no "fundamental matters" are concerned (211–214).

There are a number of problems with Urmson's arguments, most notably that they seem to apply more to club rules than to ethics. Unless we grant that morality is a matter of "codes" and that for it to be *S*'s duty to ϕ, it must be appropriate for someone or other to apply pressure to secure *S*'s compliance, (3) and (5) are beside the point. Since (4) begs the question, we are left with (1) and (2). Now, (1) and (2) are plausible; indeed, if we leave out the bit about exerting pressure, a Kantian would find them unobjectionable.[37] But, although plausible enough as claims, they offer no more support for inclusion of the category of the supererogatory in ethics than for leaving things as Kant left them. They support recognition of imperfect duties as much as they support the thesis in whose defense they are put forth. Indeed, this is the case with all of Urmson's arguments: they provide no reason (unless ordinary-language arguments are appended to them) for thinking that an ethical theory that already includes the categories of imperfect and perfect duties needs to be restructured to include the category of the supererogatory.[38]

The same holds for some of David Heyd's arguments. Heyd seeks to show "that some supererogatory acts must exist because society

[37] I here ignore a difficulty with (2): Just what does Urmson mean by "within the capacity . . ."? His attempt to clarify (2) only complicates matters; for he gives, as examples of rules that ask more than the ordinary person (or man?) is capable of, the prohibition laws in the U.S. and the gambling laws in Britain. But surely it is not beyond the capacity of the ordinary person (assuming that the ordinary person is not supposed to be an alcoholic) to abstain from liquor and dice! Urmson gets into this difficulty through his conflation of moral demands with social and legal demands.

[38] For further discussion of problems with Urmson's arguments, see Heyd, pp. 166/7; Clark, pp. 27–29; and Pybus's entire paper.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

cannot require of the individual every act that would promote the general good, and because the individual has the right to satisfy his wants and to achieve his ends and ideals regardless of their social utility (with some obvious limitations, of course)" (166).[39] We need not examine the argument for this claim; for, if successful, the claim provides as much support for the thesis that ethical theories should distinguish imperfect duties from perfect duties as it does for the supererogationist thesis that it purports to establish.[40]

But Heyd also claims that "supererogation can be proved to have moral value by pointing out the freedom of the individual involved in purely optional choice, the social cohesion resulting from supererogatory behaviour and the rationality of voluntary altruistic behaviour" (166).[41] Insofar as performance of imperfect duties is not purely optional, this claim (or part of it) would seem to provide an argument for supererogation stronger than the correlative argument for imperfect duties. It rests, however, on the controversial (and in the context of my paper, question-begging) assumption that freedom is *incompatible* with (or at least diminished by) moral constraint. If, as Kant holds, moral constraints issue from within each of us and are expressive of, rather than in conflict with, our freedom, then the fact that supererogatory acts are purely optional, whereas imperfect duties are not, is of little relevance.[42]

Heyd also suggests that social cohesion results from supererogatory behavior. A member of a group, he explains, "shows that he has an interest in his fellow members which is deeper than his contractual commitments, or than the personal benefit he can draw from his membership in the group" when he does "more than is required." This, in turn, improves the relations between the members of the group: the relations "become more friendly, personal, and based on good will" (175). Two points are in order here. First, it is clear that many acts that fulfill imperfect duties will equally express this willingness to do more than the bare moral minimum. Second, his claim

[39] Michael Clark offers a similar argument.

[40] His argument does, I believe, rest on a false dichotomy between two conceptions of a morality of duty: "a system of requirements aiming at the maximization of general good or happiness" and "a means of securing some minimal conditions of cooperation and justice" (174).

[41] I do not discuss the claim about the rationality of altruistic behavior. I find no argument or explication of it in Heyd, unless it is (despite what he says) part of an argument for the claim that a particular argument against supererogation is unsuccessful, rather than an argument that supererogation has positive moral worth.

[42] Heyd also puts forth a more plausible suggestion: "Such a freedom allows for the exercise of individual traits of character and for the expression of one's personal values and standards of moral behaviour" (175). But imperfect duties equally allow for this.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

rests on a familiar but questionable picture of obligations. We tend to assume that obligations can best (or only) be seen on a contractual and individualist model, where one regards one's obligations as a sort of burden that one must undertake in order to get what one wants. This is indeed the way that contemporary philosophy and contemporary American common-sense morality typically see obligation, but it may not be the best way. In a probing discussion of Kant's ethics, Thomas Auxter draws a striking contrast between the connotations of 'obligation' and those of the German equivalent, *Verbindlichkeit:*

> In English the sense of obligation is a "requirement that compels," "an action imposed by which one is bound or restrained," or "a constraining power." The source of this requirement, compulsion, imposition or constraint may be law, society, or conscience, but the effect is the same: one is bound (in a negative sense) to do something because one is forbidden to do otherwise. The essence of obligation is constraint. Yet the meaning of *Verbindlichkeit* is not restricted to this idea of constraint. While it is possible to speak of ourselves as being bound to do something in a negative sense (that is, as being bound, tied, and *constrained* to do something), it is also possible to think of the relation as a positive one. In this sense we are united, connected, or combined with others; we are joined and linked to them. Here the meaning of "bound" is positive. . . . Moral (that is, rational and autonomous) activity is the basis for the tie we feel with others.[43]

Thus, not only is it not the case that the category of supererogatory acts is needed instead of (or in addition to) that of imperfect duties in order to promote a sense of community; the very notion of duty may have a special contribution to a sense of community which supererogation lacks.

In sum, none of the arguments in favor of recognizing a special category is compelling. I suggest that whatever initial appeal they have is due to ignorance of or inattention to the alternatives. Compared with an ethical theory that requires that one always do as much good as one possibly can (or more generally, that regards every single good deed as mandatory, unless one could be doing something better still, in which case the better act is mandatory), Urmson's vision of a two-tiered theory is quite attractive. But, when compared with a scheme that classifies under the heading of duty acts often thought of as supererogatory and treats these as ways of fulfilling imperfect duties, the two-tiered approach loses some of its appeal.

[43] Thomas Auxter, *Kant's Moral Teleology* (Macon, Ga.: Mercer UP, 1982), pp. 163/4.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

V. REMAINING WORRIES

Even if supererogationists were convinced by my discussion that an ethical theory which subsumes under the heading of duty the acts they classify as supererogatory need not be inadequate, an issue remains. How much emphasis should be given to supererogatory acts? Should they be singled out? Kant thinks they should not be, at least not often (and certainly not as a part of moral education). The supererogationist thinks they should be.

In his chapter on Kant's ethics, Heyd offers an observation that tells in favor of the view that it is often going to be quite useful and valuable to single out supererogatory acts. He notes, "If x and y are both actions which equally fulfill a certain imperfect duty D, then it is true that we may perform *either x or y*. . . . But although Kant says nothing about it, there is the possibility of doing *both x and y*. And doing so is clearly meritorious and praiseworthy" (63).[44] Heyd's observation suggests that we need to be able to discriminate between the following two sorts of cases. Imagine Jill and Maria, both of whom have adopted a maxim of beneficence. Both, that is, really care about the welfare of others; neither is someone who just tries to fulfill a minimal requirement so that she can get on with the things she really cares about. Although the latter stance is inconsistent with having a maxim of beneficence, there are many different ways of living, all consistent with having that maxim. Jill helps others often, but she does not make the sacrifices of time and energy that Maria makes. Maria volunteers on emergency hotlines. Jill does not turn her back on needy people and is sensitive to the needs of others, but she does not go out of her way to involve herself in activities of the sort that Maria takes part in. In addition, Jill tends not to take notice of others' needs except when she is personally acquainted with the people in question (or when a crisis, such as the famine in Ethiopia, receives much more attention in her community than faraway crises usually receive).

If we do not emphasize the supererogatory, how can we pay due regard to instances in which someone does far more to fulfill an imperfect duty than do many others who themselves cannot be faulted for having too legalistic or minimalist a conception of what they ought to do? In both cases they really do have the right spirit about what they are doing; they are not trying to get duty or morality "out of the way." They take it seriously; they do not treat it the way most of us treat, say, house cleaning. And, we are assuming, neither Maria nor Jill is primarily aiming to impress others or to buttress a

[44] Michael Stocker makes a similar point in his "Supererogation and Duties," in N. Rescher, ed., *Studies in Philosophy*, APO Monograph Series, 1968.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

personal sense of moral superiority. In such a case, we surely need to have a way of according a special moral status to some of Maria's actions. What better way than by deeming her actions supererogatory?

This argument is compelling only if we are unimaginative. The strategy is not only unnecessary; it is even inadequate to meet the problem at hand. Which acts will count as supererogatory? Only the really time-consuming ones? Or perhaps those she finds most unpleasant? Or those which most of us would find especially difficult or unpleasant? None of this looks promising. Furthermore, how are we to distinguish the Marias from the Mother Theresas? These are not small, isolated problems; they are symptomatic of a general misdirection. The supererogationist approach assumes that we need a way to accord a special moral status to certain *actions*. But is this really what is needed? No. What deserve special recognition are their *characters*. It is not that some particular actions of Maria's should be characterized in a way that sets them apart from her other actions and from Jill's actions; what is ultimately impressive is the *character* that Maria's decision to work on an emergency hotline reflects.

Long before Urmson wrote "Saints and Heroes" we have had available to us another way of according special moral status to the conduct of a Maria or a Mother Theresa: we speak of their virtues, of their character, of their remarkable commitment to good causes. Not only do we have available to us this way of recognizing the special goodness of a Mother Theresa, Kant does as well. *The Doctrine of Virtue* is, after all, about virtue; it is very much in keeping with Kant's ethics to speak of a person or his character as virtuous. Kant's notion of virtue as "the strength of man's maxim in fulfilling his duty" (54 [394]), or "fortitude in relation to the forces opposing a moral attitude of will in us" (38 [380]), may seem to lend itself better to recognizing the virtue of someone who fulfills a perfect duty at enormous cost to himself than to discerning the special goodness of Maria as contrasted with Jill, or Mother Theresa as contrasted with either of them. But it is suited to the latter as well: virtue is revealed by the extent to which, and the way in which, one fulfills such duties of virtue as "participating actively in the fate of others" (126 [457]). Kant says that it is not so much a matter of "the degree to which one follows certain maxims," but rather of "the specific quality of the maxims" (65 [404]). Although Maria and Jill both adopt maxims of beneficence, their more specific maxims differ, and some of their differences reflect a differing degree of virtue.[45]

[45] I do not want to overemphasize the capability that a Kantian approach has for providing a basis for saying that Maria is more virtuous than Jill, and Mother

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

It might be objected that to do as I propose would not really be any different from emphasizing supererogatory acts. After all, will not recognition of a virtuous character be based on the judgment that the person has performed supererogatory acts? The objection, I believe, oversimplifies the relation between our appraisals of individual acts and our judgment that the agent has a virtuous character or character trait. First, although the fact that *S* performs supererogatory actions provides some evidence of virtue, it does so only in conjunction with information about *S*'s motives, aims, and values—information gleaned only from observing his nonsupererogatory actions as well as the supererogatory ones, and, more generally, from observing his conduct. So, that the agent performs supererogatory acts does not by itself provide much evidence of virtue. Second, a well-founded judgment that the agent is virtuous need not be based on a claim that he has performed supererogatory acts. The basis might be, rather, that he fulfills perfect duties at enormous cost to himself and in circumstances where most of us would not fulfill such duties. Or again, the basis might be that he does more than most people or goes to more trouble than do most people to fulfill his imperfect duties; more generally, it might be that he takes them much more seriously. A judgment that a person is virtuous may, but need not, have as part of its basis a judgment that he (often) performs supererogatory acts.

The difficulty of determining that *now*, beginning with *this* act, the agent is doing more than what is required by principles of imperfect duty, further undermines the notion that a judgment that the agent's conduct reveals generosity or deep commitment to a worthy cause should be based on the assessment that he has performed a supererogatory act. For these reasons, a judgment that someone has a certain virtuous character trait is not a disguised judgment that he has performed or often does perform supererogatory acts.

It might still be said that the judgment that someone has per-

---

Theresa yet more virtuous. A singular advantage of the Kantian approach is that it does not encourage us to rank virtuous people or virtuous lives. There are many ways of being virtuous, and though we must not be complacent and must strive to be more virtuous, it is not the case that we should have a determinate picture of the Best (exemplified by the life of a particular saint or hero) and measure ourselves accordingly.

That there are, in Kant's ethics, two obligatory ends—self-perfection and the happiness of others—may seem to allow too great a diversity of virtuous lives and too little room for saying (e.g.) that the first-rate musician (assuming that she takes the happiness of others, and not only her own perfection, seriously) is not as virtuous as Gandhi or Mother Theresa. I cannot take up this question here, nor the more general question of how much "ranking" is appropriate.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

formed a supererogatory act is in some important sense more basic or "natural" than a judgment that she is generous or in some other way virtuous. I want to suggest, however, that it may seem basic to us only because twentieth-century ethics has, until very recently, ignored character and virtue, and has attempted to say all we want to say about morality through judgments of actions. For this reason, it may be more familiar to us to locate the goodness in the act, but that it is is no argument for continuing to do so. And I see no other reason for taking the judgment that an act is supererogatory to be more basic.

I turn now to a second possible objection. One might wonder whether the reasons presented earlier for being wary of categorizing acts as supererogatory do not also apply to the view I propose.

Let us recall the problems and briefly consider whether a Kantian view gets around them. As I indicated earlier, a problem with the category of supererogatory acts is that it invites the notion that a large part of morality is beyond moral scrutiny. It does so in two different ways. First, it suggests that acts of beneficence and other acts standardly classified as supererogatory are strictly "extras," not morally mandatory. Because they are optional, self-criticism and moral scrutiny of one's omission of such acts would be inappropriate (unless, perhaps, one felt that one had a special "calling" to morality). Second, inclusion of a special category of supererogatory acts in ethics invites the assumption that admirable and apparently unselfish acts, which are more than we ordinarily expect from people, are certain to be morally unobjectionable.

The Kantian view avoids both of these problems. Concerning the first difficulty: its position is that the happiness of others is an obligatory end and that one must, therefore, embrace a maxim of beneficence, but that it is impossible to say exactly what one must do accordingly. How much one must do and precisely what, cannot be said; but one must care, and the caring must be practical, not "pathological." I must *do* something, not content myself with feeling moved. "Doing something" is not optional.[46]

The view steers clear of the second problem in the following way: if we accord a special moral status to the character of Mother Theresa rather than to her impressive actions, we avoid the tendency to suppose that actions that exhibit a remarkable and morally good character trait must be right. Acts such as those of tremendous fortitude or self-sacrifice impress us because they reveal a strength of

[46] For an excellent discussion, see Susan Mendus, "The Practical and the Pathological," *The Journal of Value Inquiry*, XIX (1985): 235–243.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

character which most of us lack. If the act is not obviously objectionable, we tend to bestow on it the label 'supererogatory'. It is more than could be asked or expected, and it displays something admirable. Yet it may nonetheless be morally questionable. If we direct our moral admiration toward the character trait rather than toward the action, this confusion can be avoided, since a character trait can be good without its being the case that any act that displays it is morally unobjectionable.[47]

In conclusion, I think that Kant was right to set aside a category of imperfect duties for those duties which are such that there is no clear answer to the question "When have I fulfilled them?" and for the fulfillment of which one must adopt a maxim. This category supplies much of what ethicists have wanted from the category of the supererogatory. The rest—except for the bare external supererogatory act, which I do not think it very important to attend to—is much better captured by evaluation of character than by recognizing a special category of actions that go "beyond" duty.[48]

What one thinks on this matter will have quite a lot to do with how broadly one wants to understand duty. I hope I have indicated that we need not start from the premise that room must be left in every ethical theory for the supererogatory, and then tailor our concept of duty accordingly.

MARCIA BARON

The University of Illinois at Urbana/Champaign
The University of Michigan at Ann Arbor

[47] This confusion is not, I think, always benign. It invites, among other things, paternalistic intervention. Kant's remarks about the need for respect to check love are apropos. See *DV*, 117–122 [449–453].

[48] The one thing that talk of virtue and imperfect duty does not capture is the bare fact that the act is beyond duty. Without the notion of the supererogatory we cannot speak of an act's being extraordinarily good except by talking about what it suggests concerning the agent's character. It is rare, I think, that we have occasion to do so. Not that there is never occasion; and so I am not urging that we seek to eliminate the term 'supererogatory' from our philosophical vocabulary. But I think everything else that we want to capture with the term 'supererogatory' is better captured by talk of imperfect duty and talk of virtue.

I should note here that supererogationists differ as to whether an act can be supererogatory regardless of the agent's motive, or whether an altruistic motive is needed before the act can qualify as supererogatory. Thus, Heyd writes that, although the intent must be altruistic, the motive need not be; it can, for example, be to gain fame (pp. 115 and 137). In contrast, Michael Clark, indicating that he uses 'meritorious' as synonymous with 'supererogatory', specifies that "the sacrifice will not be meritorious unless made from altruistic rather than merely selfish motives" (pp. 23 and 29). Supererogationists accepting the first position are particularly unlikely to find my proposal congenial, since it does not encourage attending to the bare supererogatory act.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

MORAL REALISMS AND MORAL DILEMMAS*

ONE of the most common ways to argue against the objectivity or reality of morality is to present moral dilemmas that cannot be resolved in any objective way. Many of these arguments are unconvincing.[1] Nonetheless, I will argue that moral dilemmas do show that extreme universal moral realism is false.[2] I must begin by defining both moral realism and moral dilemmas.

I. WHAT MORAL REALISM IS

Moral realism has been very popular recently, but its various defenders define it in various ways: as the claim that moral judgments must obey the law of excluded middle, as the claim that moral facts can enter into causal or explanatory relations, as the claim that moral judgments are independent of the mind, etc. Since 'moral realism' is a technical term, no such definition is the only correct one. When I define 'moral realism', I hope that my definition falls within common usage, but I do not claim that all other definitions are wrong. My aim is simply to specify the claim against which I argue.

I define 'moral realism' as a claim that moral judgments are independent of certain mental states. To say that moral judgments or their truth values depend on certain mental states is to say that, necessarily, the truth values of those judgments would change if the mental states changed in relevant ways, but other factors remained

* I am indebted to Ruth Marcus, Robert Fogelin, Bernard Gert, Willis Doney, Judith Lichtenberg, Michael Morris, Geoffrey Sayre McCord, and Mark Bedau for helpful comments on earlier drafts.

[1] The most famous argument is in Jean-Paul Sartre, "Existentialism Is a Humanism," P. Mairet, trans., in W. Kaufmann, ed., *Existentialism from Dostoevsky to Sartre* (New York: Meridian, 1957), pp. 297/8. A more precise version was given by Bernard Williams in "Ethical Consistency" and "Consistency and Realism," in *Problems of the Self* (New York: Cambridge, 1973), pp. 166–206. Williams's argument was refuted, however, by Philippa Foot in "Moral Realism and Moral Dilemma," this JOURNAL, LXXX, 7 (July 1983): 379–398. See also Ross Harrison, "Ethical Consistency," in Harrison, ed., *Rational Action* (New York: Cambridge, 1979), pp. 29–45. Williams's argument is different from mine because he considers only judgments of a single agent, whereas I compare the judgments and choices of different agents in relevantly similar moral dilemmas.

[2] My argument is derived from several predecessors: Peter Winch, "The Universalizability of Moral Judgments," in *Ethics and Action* (Boston: Routledge & Kegan Paul, 1972), pp. 151–170; Alasdair MacIntyre, "What Morality Is Not," in *Against the Self-images of the Age* (Notre Dame, Ind.: University Press, 1978), pp. 97–100; and David Wiggins, "Truth, Invention, and the Meaning of Life," *Proceedings of the British Academy*, LXII (1976): 366–372. Winch and MacIntyre direct their arguments against universalizability, but Wiggins shows that such arguments have implications for moral realism.

0022-362X/87/8405/0263$01.40
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar
© 1987 The Journal of Philosophy, Inc.

Digitized by Arya Samaj Foundation Chennai and eGangotri

the same. Moral realists deny that the truth values of moral judgments must change under such conditions.

When moral realism is defined in terms of independence, it is crucial to specify exactly what moral judgments are supposed to be independent of. The truth values of moral judgments might be independent of certain mental states but not of others (e.g., of desires, choices, and moral beliefs, but not of nonmoral beliefs or cognitive abilities), or they might be independent of the mental states of one person but not of another (e.g., the person making the judgment but not the agent who is judged),[3] or they might be independent of actual beliefs and desires but not of what someone would believe or desire under ideal conditions of some kind.[4] Furthermore, some moral judgments might be independent of such factors even though others are not.[5]

This variety of factors produces many degrees of moral realism. Moderate moral realists claim that moral judgments are independent of some but not all of the above factors. However, since I want to argue against the extreme view that moral judgments are independent of all of these factors, I need to define moral realism in an extreme way. I will call the truth conditions of a moral judgment *realistic* only if whether or not they hold is independent of the actual and ideal moral beliefs and choices of the person judged and of the person who makes the judgment. A moral judgment is then realistic just in case its truth conditions are realistic. A moral theory is a version of moral realism only if it claims that *every* moral judgment is realistic. Thus, moral realism is the claim that some moral judgments are true[6] and every moral judgment is true if and only if certain

[3] E.g., R. M. Hare [*The Language of Morals* (New York: Oxford, 1952)] claims that the acceptability of moral judgments depends on the moral principles of the judger; Williams ["Internal and External Reasons," in *Moral Luck* (New York: Cambridge, 1981), pp. 101–113] claims that the acceptability of moral judgments depends on the motivations of the judged; and Gilbert Harman [*The Nature of Morality* (New York: Oxford, 1977)] claims that the truth values of moral judgments depend on the motivations of both judger and judged. None of these theories is realistic on my definition.

[4] E.g., the truth values of moral judgments depend on ideal but not actual mental states, according to Roderick Firth in "Ethical Absolutism and the Ideal Observer," *Philosophy and Phenomenological Research*, XII, 3 (March 1952): 317–345. Contrast R. B. Brandt, *Ethical Theory* (Englewood Cliffs, N.J.: Prentice-Hall, 1959), pp. 265–269. Brandt makes ideal mental states determine whether moral judgments are justified but not whether they are true. Thus, Brandt's theory, unlike Firth's, is compatible with moral realism on my definition.

[5] E.g., David Wiggins (p. 370) claims that valuations can correspond to an independent reality, but practical judgments cannot.

[6] This condition must be added to exclude error theories like that of J. L. Mackie, *Ethics* (New York: Penguin, 1977), p. 35.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

conditions hold which are independent of the actual and ideal moral beliefs and choices of the judger and the judged. Moral realism then implies, for example, that, if abortion is morally wrong, it would still be morally wrong even if everyone who obtains or judges an abortion believed it was not wrong and, furthermore, would choose it (when possible) even if he/she were fully informed, rational, impartial, etc. Weaker theories can be called "moral realism," but the argument below will be directed against this extreme universal version of moral realism. I will return later to weaker theories to see which of them can escape the argument.

II. WHAT MORAL DILEMMAS ARE

Having defined 'moral realism', I now need to define 'moral dilemmas'. Moral dilemmas are usually defined as situations where an agent ought to adopt incompatible alternatives; but the term 'ought' is unspecific in several important ways. Consequently, I will define *moral dilemmas* as situations where there is a moral requirement for an agent to adopt each of two incompatible alternatives and where neither moral requirement overrides the other (because they are equal or incomparable).[7]

This definition might seem to beg the question against moral realism or antirealism by referring to requirements and overriding. I am using the term 'moral requirement', however, in a broad, neutral sense. A *moral reason* to adopt (or not to adopt) an alternative is or is provided by a fact that the alternative has some morally relevant property or consequence, such as that it fulfills a promise (or is unfair or a lie) or avoids causing (or causes) pain, death, or loss of freedom. A moral reason is a *requirement* if it would be morally wrong not to act on it without any moral justification or excuse. Moral realists then claim that the facts that provide moral requirements are independent of certain mental states. Moral antirealists can claim that the relevant facts are about mental states or that which facts provide moral requirements depends on certain mental states. Thus, moral realists and antirealists can both agree that there are moral requirements in my sense.

The notion of overriding is similarly neutral. One moral requirement *overrides* another only if the former is stronger over all in some morally relevant way, so that it is morally wrong to violate the former requirement. It is not enough that one of the moral requirements can somehow be ranked above the other. An agent might favor one requirement merely because doing so maximized the agent's self-in-

[7] I argued for this definition in "Moral Dilemmas and Incomparability," *American Philosophical Quarterly*, XXII, 4 (October 1985): 321–329.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

terest or merely because of a flip of a coin, but such rankings do not show that one moral requirement is stronger in a morally relevant way.[8] In order for one moral requirement to override another, there must be some ranking that has the essential features of morality, whatever they are.

Moral realists and antirealists disagree about what is essential to morality and thus about which rankings are morally relevant and about which situations are moral dilemmas. Moral realists must claim that a ranking is not morally relevant unless it is realistic, i.e., unless the higher moral requirement is so strong that anyone who violates it does what is morally wrong, regardless of his or her actual and ideal moral beliefs and choices. If this is the case, I will say that the stronger moral requirement *realistically overrides* the other. Conflicts between moral requirements that cannot be ranked realistically can be called *realistic moral dilemmas.*

In some realistic moral dilemmas, moral antirealists might still claim that there is a personal moral ranking, i.e., that one and only one moral requirement is such that it would be morally wrong for the individual agent to violate it, even though similar acts would not be wrong for all other agents. If there is a personal moral ranking, the situation is not a personal moral dilemma, and moral antirealists might refuse to call the situation a "moral dilemma" at all. Nonetheless, such situations can still be realistic moral dilemmas, and moral realists might insist that they are moral dilemmas. Whether or not they are called "moral dilemmas," it is such conflicts between moral requirements that can be ranked personally but not realistically which cause trouble for moral realism. I will count such situations as moral dilemmas so that I can ask simply whether moral realism can handle moral dilemmas.

Thus, moral dilemmas can be defined as situations where there is a moral requirement for an agent to adopt each of incompatible alternatives, but where neither moral requirement is overridden in any way that is both morally relevant and realistic. This definition may not capture common usage, if there is any, but it does pick out the kind of situation that I will discuss.

III. THE ARGUMENT

Given these definitions, I can now present my argument against extreme moral realism from moral dilemmas. My strategy will be to compare moral judgments of different agents in relevantly similar moral dilemmas. The basic idea is that realistic moral facts do not

[8] Cf. Ruth Barcan Marcus, "Moral Dilemmas and Consistency," this JOURNAL, LXXVII, 3 (March 1980): 121–136, pp. 135/6.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

favor either alternative in a moral dilemma, but different agents can still personally favor different alternatives, and then their personal rankings or choices can determine what they ought to do. Since moral realists deny that any moral judgments depend on such mental factors as moral beliefs or choices, extreme universal moral realism is false.

The argument is clearer if we look at an example of a moral dilemma. Suppose Jim promised a professional colleague to finish a project by a certain date, but, through no fault of his own, he is late and has only one day left. On the other hand, today is his daughter's birthday, and the family usually goes sailing on her birthday. Jim did not promise to go, but he is expected to come along, and his daughter will be disappointed if he does not. Thus, Jim cannot keep his professional promise without causing some pain to his daughter. I assume that there is a moral requirement to keep the promise and a moral requirement not to cause the pain, but the importance of the promise and the amount of pain are balanced so that neither requirement realistically overrides the other. The promise is not so important that it would be morally wrong for anyone to break it in a relevantly similar situation. Also, the pain is not so great that it would be morally wrong for anyone to cause it in a relevantly similar situation. The conflicting moral requirements might be described as equal or incomparable, but Jim's situation is a realistic moral dilemma in either case.[9]

Now suppose that Jack is in a moral dilemma that is similar to Jim's in all respects that could be morally relevant according to moral realism. Jack's keeping his promise would cause just as much pain to Jack's family as Jim's keeping his promise would cause to Jim's family. Jack's promise is just as important professionally as Jim's. Thus, just as in Jim's case, neither moral requirement realistically overrides the other in Jack's case.

Despite such realistic similarities, Jim and Jack still personally rank the moral requirements differently. Jim believes that his daughter's pain is more important to him than his professional promise, because he chooses to be a family man committed to a family way of life. In contrast, Jack personally ranks his professional duty and promise more highly than the moral requirement not to cause pain to his daughter, because he chooses to be a professional man committed to

[9] A critic might deny that moral dilemmas are possible, but I argue for the possibility of moral dilemmas in "Moral Dilemmas and Incomparability;" and I respond to counterarguments in " 'Ought' Conversationally Implies 'Can'," *Philosophical Review*, XCIII, 2 (April 1984): 249–261; and in "Moral Dilemmas and 'Ought and Ought Not'," *Canadian Journal of Philosophy*, forthcoming.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the professional way of life. Jack and Jim consider exactly the same realistic factors, and both consider only factors that both agree are morally relevant. Neither person's choice or personal ranking depends on any error, ignorance, irrationality, or partiality. Neither way of life is morally wrong, since neither person is so committed to his way of life that he would violate a realistically overriding moral requirement in order to further his family or profession. Nonetheless, Jim and Jack arrive at different personal rankings of the moral requirements.

The crucial question for moral realism is: Which moral judgments are true in this situation? I will discuss judgments about what each agent ought to do, and I will distinguish different interpretations of these judgments. Nonetheless, my conclusions can be extended to other moral judgments.

The argument against extreme moral realism now runs as follows:

(1) Jack ought to keep his promise.
(2) Jim ought not to keep his promise.[10]
(3) If Jim ought not to keep his promise, then it is not the case that Jim ought to keep his promise.
(4) Therefore, it is not the case that Jim ought to keep his promise.
(5) Therefore, whether or not an agent ought to keep his promise depends on the agent's choice and personal ranking.
(6) If moral realism were true, what an agent ought to do would not depend on the agent's choice or personal ranking.
(7) Therefore, moral realism is false.

(1) and (2) are the judgments that cause trouble for moral realism. (3) is an assumption that is needed to derive (4). (5) follows from (1), (4), and the supposition that there is no relevant difference between Jack and Jim other than their personal rankings and choices. (6) is supposed to follow from the definition of 'moral realism'. (7) follows from (5) and (6).

The crucial premises of the argument are then (1), (2), and (3). It might seem easy for a moral realist to escape simply by denying (1) or (2), i.e., by denying either that Jack ought to keep his promise or that Jim ought not to keep his promise (or both). However, whether this denial is plausible depends on how (1) and (2) are interpreted; so we need to consider various interpretations of the term 'ought'. The

[10] A moral realist might respond that it is not really true that Jim ought not to keep his promise, but only that he ought not to hurt his daughter. However, I assume here and throughout that 'ought' is extensional enough that an agent ought not to do what is sufficient for something (else) that he ought not to do. Jim cannot keep his promise without hurting his daughter. Thus, if Jim ought not to hurt his daughter, he ought not to keep his promise.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

argument is sound only if there is a single interpretation of 'ought' on which (1), (2), and (3) are all true.

IV. REALISTIC INTERPRETATIONS

There are many interpretations of 'ought' on which the argument fails. One interpretation is that an agent ought to do that for which there is a moral reason that is not realistically overridden. On this interpretation, (1) and (2) follow from the supposition that Jack and Jim are in moral dilemmas; so moral realists cannot deny these premises. However, premise (3) then claims that, if Jim has a non-realistically-overridden moral reason not to keep his promise, he cannot also have a non-realistically-overridden moral reason to keep his promise. This is false, because Jim is in a moral dilemma, so he has non-realistically-overridden moral reasons both to keep and not to keep his promise. Since the same judgments are true of Jack and of any other agent in a similar dilemma, this interpretation of 'ought' causes no trouble for moral realism. This response can be extended to any interpretation where both 'ought' and 'ought not' can be true. On any such interpretation, (3) is false; so the argument fails.

A stronger interpretation of 'ought' is that an agent ought to do only that for which there are realistically overriding moral reasons. On this interpretation, premise (3) is true, because, if there are realistically overriding moral reasons to do something, there cannot also be realistically overriding moral reasons not to do the same thing. However, premises (1) and (2) are not true on this interpretation. (1) now claims that Jack has a realistically overriding moral reason to keep his promise, but this cannot be true, since Jack is in a moral dilemma. (2) is false for similar reasons, so the argument fails to refute moral realism. Similar responses apply to all other realistic interpretations that make 'ought' contrary to 'ought not'.

Thus, on some realistic interpretations, both Jack and Jim both ought and ought not to keep their promises, but, on other realistic interpretations, both Jack and Jim neither ought nor ought not to keep their promises. If such different interpretations are confused, the above argument might seem to refute moral realism. However, on each realistic interpretation, the same moral judgments are true of both agents; so their truth values need not depend on any subjective difference between the agents. Consequently, the possibility of moral dilemmas is compatible with realistic truth conditions for at least some moral judgments: judgments about what there is a moral reason or requirement to do; judgments about which moral reasons or requirements realistically override which; and judgments where 'ought' is interpreted in terms of moral reasons and requirements and their relative realistic strengths. These are the only moral judg-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ments that are essential for moral dilemmas, as moral realists define them; so moral realists can consistently accept the possibility of moral dilemmas, as they see them.

V. ANTIREALISTIC INTERPRETATIONS

This still does not settle the issue, because realistic interpretations miss the point of the antirealist argument. The point is that moral dilemmas create a need for another kind of judgment that does more than merely describe moral requirements and their relative realistic strengths. Since neither moral requirement realistically overrides the other, no realistic moral judgment can help the agent choose. But the agent must and can choose, and the choice need not be arbitrary, since it can be based on personal factors. When an agent has chosen on a personal basis, the agent often says "I ought to do this" or "This is what I ought to do" or even "I have decided that I ought to do this." It is such judgments with 'ought' that cause problems for moral realism. Since they cannot depend on any realistic factor, they must depend on some nonrealistic factor.

Some moral antirealists interpret such judgments by claiming that they are used merely to *express* a decision or a commitment.[11] This is often true, but it is not enough to refute moral realism. Moral realists can admit that moral judgments are often used for speech acts such as to express decisions and commitments; they deny only that the truth values of moral judgments depend on decisions and personal commitments. Thus, the argument against moral realism would confuse speech acts with truth conditions if it depended on any such interpretation of 'ought'.

In order to refute moral realism, judgments with 'ought' must be interpreted so that their truth values depend on some subjective factor. But not just any subjective factor will do. Some agents choose without any reason at all, but then it is not true that they *ought* to do as they choose. Other agents choose on the basis of self-interest, and then they prudentially ought to do as they choose; but it is still not true that they *morally* ought to do as they choose. Such cases cause no trouble for moral realism.

The troubling cases are when an agent chooses an alternative because of his commitment to what can be called "a way of life." This notion is vague, but what is important here is that a way of life includes a tendency to rank one kind of value above another and to choose accordingly. Such rankings are moral beliefs, and such ways of life can be chosen. But moral realists deny that such mental factors as choices and moral beliefs affect the truth values of moral judg-

[11] Such speech acts are emphasized in Winch, p. 165, and MacIntyre, p. 98.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ments. Thus, moral realists must deny that the truth values of moral judgments depend on the way of life of the judged agent.

However, such ways of life *do* affect what agents ought to do in some moral dilemmas. If Jim asked me what he ought to do, I would tell him that he ought to break his promise in order to avoid hurting his daughter. Why? Because this choice is required by his commitment to a way of life that is morally and rationally permissible, so to choose otherwise would show a lack of integrity on his part. Jim's family way of life is rationally permissible, because his choice of it does not depend on any defect such as error, ignorance, irrationality, or partiality. Jim's family way of life is also morally permissible, because he is not so committed that he would help his family if there were a realistically overriding moral requirement not to do so. Still, if he hurt his family when there was no realistically overriding moral reason to do so, he would violate his way of life and his integrity.

Furthermore, the agent's commitments and personal rankings are the proper basis for choices in situations of this kind. Since the situation is a moral dilemma, the choice cannot be based on realistic moral factors or on other people's interests, since these are balanced. The personal rankings of the people affected cannot be used to decide, since several people are affected and their rankings might conflict. It might seem that the person making the judgment can use her own ranking, but there is no justification for imposing a ranking on an agent when the agent can reject that ranking without any rational or moral defect. The judger might not even be affected by the choice, but the agent is responsible for the decision; so it is the agent's own ranking and way of life that determine what he ought to do. Therefore, even if the judger is committed to the professional way of life, it is still true that Jim ought not to hurt his daughter.[12]

In contrast, if Jack asked me what he ought to do, I would tell him that he ought to keep his professional promise and to hurt his daughter. This different judgment is true, because Jack is committed to a different way of life, a professional way of life which is also morally and rationally permissible. If he violated his professional obligation when he had no realistically overriding moral reason to do so, this would show a lack of integrity on Jack's part. Thus, Jack ought to do what Jim ought not to do, and vice versa, even though their situations are similar in all respects that can be morally relevant according to moral realism.

[12] It is not clear whether Jim must or is required to break his promise, since it is not clear whether a choice contrary to his way of life would make him liable to punishment in one of its many forms. In any case, my argument does not require this stronger judgment.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

In order to interpret such judgments, moral antirealists need to build ways of life, choices, or personal rankings into the truth conditions of moral judgments. There are many ways to do this, but let me suggest one: An agent ought to do something if (and only if) there is a moral reason for the agent to do it, this reason is not overridden realistically, and the agent personally ranks this reason higher than any conflicting moral reason because it fits best into a rationally and morally permissible way of life that the agent chooses without depending on any defect such as error, ignorance, irrationality, or partiality. This interpretation does not capture everyone's use of 'ought', but it is coherent and does capture the point of the antirealist argument.

On this interpretation, Jack ought to keep his promise and to hurt his daughter, but Jim ought not to keep his promise or to hurt his daughter; so (1) and (2) are true. On this interpretation (3) is also true, since, if keeping his promise fits best into a way of life, breaking the same promise in the same situation cannot also fit best into the same way of life. The argument thus shows that, on such interpretations, judgments with 'ought' do not have realistic truth conditions. Therefore, extreme universal moral realism is false.

VI. RESPONSES

Several objections might be raised to this argument. First, moral realists might respond that the antirealist interpretation of 'ought' fails to capture common usage, but moral realism is a claim about common usage; so the antirealist interpretation fails to refute moral realism. However, this response either overestimates what antirealists need to claim or underestimates the flexibility of common usage. The argument is not supposed to show that no moral judgments are realistic; so moral antirealists need not claim that their interpretation captures all uses of 'ought'. And the antirealist interpretation does capture some uses of 'ought' that do not violate any rules of common usage. It is not unintelligible to claim that an agent's way of life affects what that agent ought to do, at least in special cases such as moral dilemmas. Many competent English speakers make or assume such judgments.

Moral realists might try to explain such uses in terms of speech acts and pragmatic effects rather than truth conditions. No pragmatic explanation, however, seems to be adequate, because it seems true that Jack ought to keep his promise but not that Jim ought to keep his promise, regardless of which speech act is done or which pragmatic effect is intended. This suggests that the agents' ways of life determine the truth values of the judgments rather than only their speech acts and pragmatic effects. Thus, the first response fails.

A second response might be to deny that, on the antirealist inter-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pretation, judgments with 'ought' are *moral* judgments (see Marcus, *op. cit.*, pp. 135/6). If such judgments are not moral judgments, then moral realists can admit that they lack realistic truth conditions but still claim nonetheless that all moral judgments have realistic truth conditions. Premise (6) of the above argument is then false; so the argument fails to refute moral realism.

It must be admitted that many judgments with 'ought' are not moral judgments. For example, if someone judges that Jim ought to break his promise, but only because breaking his promise is in his interest, then this judgment with 'ought' is not a moral judgment. I specified above, however, that Jack and Jim do not consider any nonmoral factors, and I also assume that the judger considers only moral factors and the agents' rankings of them.

If so, there is no good reason to deny that judgments with 'ought' on the antirealist interpretation are moral judgments. Such judgments have all the essential features of moral judgments. Regarding content, they concern harm to others. Regarding force, they can override nonmoral judgments. Regarding form, the tricky question is whether judgments with 'ought' on the antirealist interpretation are universalizable. They are in one way but not in another. These judgments are universalizable in the sense that all agents who choose the same way of life ought to do the same acts in relevantly similar situations. They are not universalizable in the stronger sense that similar judgments are true of all agents in relevantly similar situations, since not all agents choose the same way of life. However, it begs the question to assume that moral judgments must be universalizable in this strong sense. This definition of morality assumes that differences in choices and ways of life are never morally relevant—but that is precisely the issue between moral realists and antirealists. Furthermore, if moral realists define moral judgments so that no judgment is moral unless it is realistic, it becomes circular for them to claim that all moral judgments are realistic (cf. MacIntyre, *op. cit.*, 99). This claim does not become completely empty, since moral realists still claim that some judgments are both realistically true and moral. Nonetheless, the universality of moral realism—the claim that not only some but all moral judgments are realistic—rests solely on a definition. And there is still no non-question-begging justification for excluding the antirealistic judgments from the domain of morality. In the absence of any such justification, I conclude that these antirealistic judgments are moral judgments; so the second response fails.

As a third response, moral realists might argue that, even though different moral judgments are true of the two agents, there must be some morally relevant difference between the agents if they are

CC-0. In Public Domain. Panini Kanya Maha Vidyalaya Collection.

Digitized by Arya Samaj Foundation Chennai and eGangotri

committed to different ways of life. Similar actions by different agents can cause different amounts of harm when the agents are committed to different ways of life. If Jim hurts his daughter, his self-esteem and his daughter's trust will be damaged. Some similar effects might flow from Jack's hurting his daughter, but the harm will be less, since he will not have violated his way of life, and, if his daughter knows this, she should expect and understand such treatment. Similarly, if Jack breaks his promise, he will lose more self-esteem and professional trust than Jim will lose if he breaks his, since Jack is committed to the professional way of life, but Jim is not. A moral realist can cite such differences in order to explain why Jack ought to keep his promise but it is not true that Jim ought to keep his.

This response, however, avoids the problem without solving it. Even if the agent's way of life does affect how much harm each alternative would cause, the situations can still be redescribed as moral dilemmas, simply by adding extra harms onto one alternative to balance the harm caused by violating the agent's way of life. For example, suppose Jim made his promise to more of his professional colleagues, so there will be more professional loss if he breaks it. Jim's breaking his promise can then cause just as much harm as Jack's breaking his, even though it is not supported as strongly by Jim's way of life. Similarly, suppose Jack has more children who will be hurt along with his daughter if he keeps his professional promise, so that Jack's promise breaking prevents just as much harm as Jim's does, even though Jim's way of life supports his promise breaking more strongly. With some such changes, the example can be described so that neither moral requirement realistically overrides the other, even though the way of life of each agent supports one alternative above the other. The agents are then in realistic moral dilemmas, and the argument goes through as before.

A final response is that moral realists do not deny that the truth values of moral judgments can depend on past choices, since they admit that whether an act is morally wrong can depend on whether the agent chose to promise to do it.[13] Promises, however, are not like commitments to ways of life. Promises are speech acts; so one can choose to make a promise but never make or even be in a position to make that promise, and then no obligation is created. And a promise does create an obligation even if there is no subjective intention to fulfill it. Thus, promises themselves, rather than choices or any other

[13] Geoff Sayre McCord raised this objection in correspondence. Similar points apply to obligations that arise from decisions to have children, etc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

subjective accompaniment, are what make such moral judgments true for a moral realist. In contrast, Jim's and Jack's subjective commitments determine what they ought to do, on the anti-realist interpretation, even if their commitments are not reflected in any objective actions or speech prior to this moral dilemma. Furthermore, personal rankings are moral beliefs, but moral realists deny the claim that whether or not one has a moral obligation to keep a promise depends on one's moral beliefs. Thus, the argument shows that the truth values of moral judgments in moral dilemmas depend on more subjective factors than in standard cases of promising. If moral realists admit that the truth values of moral judgments can depend on such subjective factors as the way of life chosen by the agent, then they grant all that the argument attempts to show.

Moral realists might have other responses, and the argument does rely on some intuitions that might be questioned; so the argument is not a conclusive proof. Nonetheless, in the absence of any adequate response, I conclude that some moral judgments do not have realistic truth conditions, and so extreme universal moral realism is false.

VII. QUALIFICATIONS

It is important to realize what this argument does not claim to show. The argument does not try to prove that every version of moral realism is false. It is aimed at only the most extreme version. Less extreme versions of moral realism are not affected by the argument.

First, the argument is directed against only universal moral realism, not partial moral realism. The argument shows only that some moral judgments lack realistic truth conditions, but other moral judgments still might have realistic truth conditions. The argument does raise the problem of where to draw the line, and it might shift the burden of proof. Still, the argument does not directly attack the claim that some, but not all, moral judgments have realistic truth conditions. And that is all that some moral realists wanted to claim anyway.

Second, even for those moral judgments which the argument shows not to be realistic, the argument refutes only the extreme claim that their truth conditions are independent of even ideal mental states of agents. It is not plausible to make judgments with 'ought' depend on the agent's choice, if the agent would choose differently if he or she were informed or rational. Thus, judgments with 'ought' depend at most on ideal, not actual, choices, so the argument does not refute a moderate moral realist (or objectivist) who claims independence from actual, but not ideal, mental states.

Third, the argument shows dependence on the choices and beliefs of only the agent, not the judger. On the antirealist interpretation,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

anyone can judge that Jim ought to break his promise and that Jack ought to keep his promise. The truth of these judgments does not depend on who makes the judgment.[14] Thus, the argument can be accepted by moderate moral realists who claim the truth values of moral judgments are independent of only the judger's mental states. This is enough to make moral judgments parallel to some scientific judgments, such as psychological judgments, so some moderate moral realists might be content with this much independence.

These qualifications should allay the fears of opponents who think that my argument leads to an unacceptable form of subjectivism or relativism. If extreme moral realism were the only alternative to moral relativism, it would be disturbing to find that extreme moral realism is false. However, my argument applies only to some moral judgments and only in moral dilemmas, so it does not lead all the way to wholesale relativism.

This might seem to take the sting out of my conclusion. However, it is important to realize that moral realism has its limits, because these limits make moral realism more plausible. Extreme moral realism underestimates the complexity of moral life when it overlooks the importance of individual choices and moral beliefs in moral dilemmas. More limited moral realism can recognize that such subjective factors are morally relevant in moral dilemmas, and it can still impose realistic limits on choices so as to avoid the problems of wholesale relativism. Thus, limited moral realism can accept the insights and avoid the problems of more extreme theories.

Nonetheless, I have not argued that limited moral realism is true. There are several other arguments against moral realism, and some apply to limited moral realism as well as to extreme moral realism. These other arguments would have to be refuted in order to defend even limited moral realism. I am not sure whether this can be done. In any case, all I have tried to do here is to show that the only version of moral realism that might be defensible is limited moral realism.

WALTER SINNOTT-ARMSTRONG

Dartmouth College

[14] Contrast Winch (p. 153) and MacIntyre (p. 98), who claim that their antirealist judgments can be made only in the first person, since their interpretations emphasize speech acts that can be done only in the first person. Cf. fn. 11.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS AND CRITICISM

SCHEFFLER ON THE INDEPENDENCE OF AGENT-CENTERED PREROGATIVES FROM AGENT-CENTERED RESTRICTIONS

IN his recent book, *The Rejection of Consequentialism,** Samuel Scheffler argues in favor of departing from purely consequentialist moral conceptions and incorporating into such conceptions what he calls an *agent-centered prerogative.* The agent-centered prerogative, unlike pure consequentialism, permits agents to refrain from always attempting to produce the best over-all state of affairs (5). At the same time, Scheffler argues against the deontological view that attempting to produce the best over-all state of affairs may sometimes be morally wrong—what he calls "agent-centered restrictions" (2, 5). In other words, Scheffler defends a position that always *permits* an agent to try to produce the best over-all state of affairs, but does not always *require* the agent to do so.

Scheffler's argument for an agent-centered prerogative is based on the now standard criticism of consequentialism, namely that consequentialism violates personal integrity.[1] A consequentialist may not give to his own projects any more weight in the moral calculus than to others' projects, just because they are his. Yet just such a view of the priority of one's own projects seems to be required for personal integrity (57, 61/2). Moreover, even if the consequentialist treats personal integrity as a good, he cannot give *his* integrity more weight than the integrity of others (61).[2] The agent-centered prerogative is Scheffler's attempt to make moral space for the priority of one's own projects because they are one's own.

On the other hand, Scheffler rejects the deontologist's answer to consequentialism, that is, agent-centered restrictions. Concern for an individual's point of view regarding her own projects cannot explain why it is categorically wrong to engage in certain acts, especially if not engaging in those acts is on balance destructive of some persons' projects. Indeed, it is difficult to see how any act could be

* New York: Oxford, 1982; reviewed in this JOURNAL by Stephen Darwall, LXXXI, 4 (April 1984): 220–226. Parenthetical page references to Scheffler will be to this book.

[1] Bernard Williams, "A Critique of Utilitarianism," in J. J. C. Smart and Williams, *Utilitarianism: For and Against* (New York: Cambridge, 1973), pp. 116/7.

[2] See also Judith Lichtenberg, "The Right, the All Right, and the Good," *Yale Law Journal,* XCII (1983): 544, 549. Of course, the consequentialist may rejoin that, if integrity is important at all, it is irrational not to seek to maximize it to the greatest extent possible. See David Mapel, Book Review, *Political Theory* (August 1983): 469/70. See also Lichtenberg, *op. cit.,* at p. 558.

0022-362X/87/8405/0277$00.70 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

categorically wrong if its nonperformance would result in even more acts of its kind.[3] Thus, the rejection of consequentialism leads Scheffler to his agent-centered prerogative, but not to the deontologist's agent-centered restriction.

To hold this position, Scheffler must argue for what he calls the *independence thesis,* namely, that rejection of agent-centered restrictions does not entail rejection of the agent-centered prerogative (81–83, 91/2, 100/1). Suppose that an agent A could sacrifice one victim V_1 and thereby save five other victims V_2–V_6 from the same fate as V_1's or worse. The best over-all state of affairs is the sacrifice of V_1 with the consequent saving of V_2–V_6. The deontologist, with his agent-centered restrictions, might say that, despite its producing the best consequences over all, A's sacrifice of V_1 is not permitted; that is, is morally wrong. The consequentialist, who rejects both agent-centered restrictions and the agent-centered prerogative, would respond that A's *failure to sacrifice* V_1 is not permitted; that is, is morally wrong. Scheffler argues, however, that both the deontologist and the consequentialist are mistaken: A may sacrifice V_1 (no agent-centered restriction), but may also not sacrifice V_1 (the agent-centered prerogative).

The difficulty with Scheffler's "hybrid" moral conception can easily be illustrated by reference to the previous hypothetical. Because of the agent-centered prerogative, V_1 may refuse to sacrifice himself for V_2–V_6. That is, V_1, like A, is not required to produce the best over-all state of affairs. But, without an agent-centered restriction on A, A *may* (though not must) sacrifice V_1, even without V_1's consent, to save V_2–V_6. And, if A may sacrifice V_1 to save V_2–V_6, it would appear to follow that A may *coerce* V_1 to *act* for the sake of V_2–V_6. Yet, if V_1 may be coerced to act to save V_2–V_6, it must be because V_1 has *no right not to act* to save V_2–V_6. The question then is: If V_1 has no right not to act to save V_2–V_6, in what way is he permitted, through the agent-centered prerogative, not to act to save V_2–V_6? It appears that the agent-centered prerogative may require an agent-centered restriction that is exactly coterminous with the prerogative in order to have any real value to the agent. In other words, rejection of thoroughgoing consequentialism should lead to deontology, not to Scheffler's hybrid position.

One might initially object to the link I have forged between A's *sacrificing* V_1 for A's consequentialist ends and A's *coercing* V_1 to act for those ends.[4] Thus, one could argue that V_1's prerogative

[3] Scheffler, p. 82. But see Lichtenberg, *op. cit.*, pp. 557–559.
[4] I owe this point to Simon Blackburn.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

creates a side constraint on A's consequentialism, a constraint that precludes coercing V_1 but does not preclude sacrificing V_1's interest in other ways. Such a view would face tough sledding, however, for it is not at all clear why one would wish to assert a right against being coerced (by, say, threats of harm) to act to aid V_2–V_6, but to reject a right against being, say, killed to aid V_2–V_6. It is a strange agent-centered prerogative that shields me from being thrown into jail for a few months for not laboring on behalf of the poor, but allows me to be hurled into the path of a boulder to prevent harm to others. It is much more plausible to assume, as I have done, that whatever can be said in favor of a right not to be coerced can be said in favor of a right not to be sacrificed.[5]

Moreover, the argument that there is a right not to be coerced to act for consequentialist ends but no right not to be sacrificed for those ends is an argument unavailable to Scheffler. That is so because the argument assumes the existence of at least one agent-centered restriction—the right not to be coerced to act for consequentialist ends—whereas Scheffler rejects all agent-centered restrictions.

In any event, Scheffler does not attempt to defend his thesis by severing the link between sacrificing and coercing. Scheffler admits that A may force V_1 to do something that V_1 need not do voluntarily (37). But he does not view this as rendering his position unintelligible or inconsistent.[6] The intelligibility of his position can be defended as follows:[7] There is a distinction between a moral (or legal) permission to do A and a moral (or legal) right to do A. The right entails a duty on others to refrain from coercing you not to do A. The permission, however, is quite consistent with a permission to others so to coerce you. We are morally permitted to do many more things than we have moral rights to do. Scheffler's is an argument for anticonsequentialist moral permissions, not moral rights. Indeed, he rejects moral rights. Thus, the fact that his anticonsequentialist moral permissions conflict with (equally valid) pro-consequentialist moral permissions does nothing to discredit his argument. Moral permissions, unlike moral rights, can logically conflict, and sometimes do so.

[5] Of course, even for the thoroughgoing consequentialist there are some acts, like inflicting pain or death on oneself or one's loved ones, which are so difficult for the agent or run so against the consequence-justified grain, that we would *excuse* her from performing them, though we would not deem their non-performance morally *justified*. With respect to such acts, we might rule out coercion by threats, but not other forms of sacrifice, *on purely consequentialist grounds*.

[6] Correspondence with the author, February 24, 1983.

[7] I owe this formulation to an anonymous editor of this JOURNAL.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

This defense of Scheffler, although it demonstrates that his argument that the agent-centered prerogative does not entail agent-centered restrictions is not incoherent, fails to render at all plausible his *normative* case for an agent-centered prerogative without any agent-centered restrictions. First, normally, where we encounter conflicting moral permissions, they are very narrow in scope and are tightly hemmed in by moral rights (agent-centered restrictions). For example, in a competitive sport such as football, the runner has a permission to attempt to advance the ball, while the defense has a conflicting permission to attempt to prevent him from doing so. In the realm of free enterprise, one competitor may have a permission to attempt to draw customers away from another's business, while the latter has a permission to attempt to thwart the former's attempt. Nonetheless, these conflicting permissions—which usually are found in areas of life where there is either a consequentialist or a deontological justification for processes of competition and a normal human motivation to act competitively if permitted—are quite limited in character. The defense may tackle the runner; it may not shoot him. A business competitor may lower her prices; she may not burn competing businesses or shanghai customers.

Even outside the area of beneficial competition, it may be plausible in certain circumstances to posit conflicting moral permissions. Nancy Davis[8] suggests that, where two people pose *passive threats* to each other, each may have a moral permission to attempt to eliminate the threat to herself by taking action that maintains or enhances the threat to the other. This moral permission is an agent-relative permission and does not provide a justification for third-party intervention on behalf of either side (*ibid.*, 191–196).[9]

Davis's agent-relative permission with respect to passive threats is one example of Scheffler's more general agent-centered prerogative. Agent-relative permissions fall in between ordinary moral excuses and ordinary moral justifications in that they are stronger than excuses—they can be invoked by the fully rational, knowledgeable,

[8] "Abortion and Self-Defense," *Philosophy and Public Affairs*, XIII, 3 (Summer 1984): 175–207, pp. 190–192. Davis gives as an example of a passive threat a mountain climber, Ben, who, because of a rockslide, is about to fall on a ledge occupied by his companion Alice. If he falls on the ledge, he will not be seriously hurt, but Alice will be killed. Alice, says Davis, is morally permitted to manipulate Ben's rope to cause him to miss her ledge and fall to his death. Similarly, Ben is morally permitted to resist Alice's efforts.

[9] See also Thomas Nagel, "The Limits of Objectivity," in Sterling McMurrin, ed., *The Tanner Lectures on Human Values*, vol. I., (New York: Cambridge, 1980), pp. 77–139.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and controlled—but, unlike justifications, they do not extend to third parties.

Davis limits her case for agent-relative permissions to situations of self-defense. She implies, moreover, that the permission to kill in order to preserve one's own life does not extend beyond permission to eliminate passive threats. It does not include, for example, permission to kill persons who merely possess something—perhaps an organ or a vaccine—which is necessary for one's own survival (Davis, *op. cit.*, 194). Even so limited, Davis's argument is troublesome and problematic.

But Scheffler's agent-centered prerogative—which is not limited to cases of preserving life or hemmed in by agent-centered restrictions that would rule out, say, deliberately harming others—is far more troublesome. As Shelly Kagan[10] points out, if Scheffler's agent-centered prerogative permits us to refuse to pay $10,000 to save another's life, then, given the absence of agent-centered restrictions, it should permit us to kill another to gain $10,000.

Moreover, consider the following theoretical problems that Scheffler's theory raises: If A can coerce V_1 to act for V_2–V_6, though V_1 has no moral duty to do so, may V_1 use self-defense to resist A's coercion? Could someone then come to A's aid to help A coerce V_1? Could someone else come to V_1's aid to help him resist A's coercion? These permissions—which are tantamount to moral anarchy—all appear to follow from the simultaneous acceptance of an agent-centered prerogative and rejection of agent-centered restrictions.

Now Scheffler has a response to the moral-anarchy objection. Scheffler's theory is not a theory that consists entirely of an agent-centered prerogative. Rather, it contains restrictions, moral prohibitions, which are consequentialist-justified rather than deontological. The agent-centered prerogative follows from permitting agents to attach more weight to their projects than those projects receive in the consequentialist calculus, but not infinitely more weight (20). At some point the consequentialist considerations override the prerogative and limit it, though Scheffler spends no time exploring this necessary limit.

Given that his agent-centered prerogatives are limited by consequentialist considerations, Scheffler might respond to the moral-anarchy scenario by pointing out that the initial conflicting permissions to A, V_1, and their allies at some point threaten consequences sufficiently suboptimal so that those permissions, or at least some of

[10] "Does Consequentialism Demand Too Much?", *Philosophy and Public Affairs*, XIII, 3 (Summer 1984): 239–254, p. 251.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

them, will be withdrawn. This response is adequate to meet the moral-anarchy objection, but it comes at far too high a price, for two reasons.

First, the response exposes the precarious status of the agent-centered prerogative. A moral agent may have an agent-centered prerogative to do *A* only if another moral agent does not exercise *her* agent-centered prerogative to do *B*. Of course, within consequentialist moral theories other than certain (implausible) types of rule-consequentialist theories, the permissibility or obligatoriness of an act is always dependent upon how others are acting. Usually, however, it is possible theoretically for properly motivated actors to coordinate their acts at an optimal level.[11] Given Scheffler's agent-centered prerogative—his enclave of discretion to deviate from producing optimal consequences—no stable coordination is morally obligatory and thus realizable (except fortuitously) among properly motivated actors.

Second, there is not even available to Scheffler a stable floor, a consequentialist baseline, a state of affairs that is the worst such state an actor is permitted to produce. For any actor facing such a state of affairs would, in the absence of agent-centered restrictions, presumably still be able to weight her projects more than they are weighted in the measurement of the baseline state of affairs. She may then presumably act to produce an even worse set of consequences, which in turn permits other actors themselves to produce still worse consequences, ad infinitum. Even if each actor can give her projects only slightly more weight than they are accorded in the impartial consequentialist calculus, the absence of agent-centered restrictions results in the possibility of an endless downward spiral of worse and worse consequences.

An equally important reason why Scheffler cannot plausibly posit conflicting moral permissions across the board, rather than as limited to discrete spheres such as athletic and business competition and (more problematically) self-defense, is that, without agent-centered restrictions, there is one important actor that is always obligated to produce an optimal set of consequences. That actor is the government. The government is not the kind of moral agent which can possess an agent-centered prerogative with respect to its own acts. It must always act as a thoroughgoing consequentialist, giving only impartial consideration to individuals' weightings of their own projects.

[11] See Donald Regan, *Utilitarianism and Cooperation* (New York: Oxford, 1980).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Now Scheffler's positing of an agent-centered prerogative has absolutely no implications for how government must act. Because it has no such implications, it also places the government in direct opposition to all exercises of Scheffler's agent-centered prerogatives. To the extent that the agent-centered prerogative entails moral permissions beyond those which are consequentialist-justified, to that extent it necessarily conflicts with what the state is morally required to see happen. The state would be obligated to attempt to prevent all non-consequentialist-justified exercises of the agent-centered prerogative, perhaps even to the extent of denying the existence of such a prerogative and inculcating a purely consequentialist morality. Moreover, because resisting the government might have dire consequences where the government is acting as a proper consequentialist and where resistance would be contagious, the consequentialist limit on the agent-centered prerogative might dictate that the prerogative never be exercised. What is novel in Scheffler's theory—the agent-centered prerogative—has absolutely no relevance to arguments about the proper use of government coercion, and perhaps no relevance either to moral education or to action under a good consequentialist government.

I conclude that, although both consequentialism and deontology have their difficulties, both in their pure forms and when joined with each other, Scheffler's mix of consequentialism and the agent-centered prerogative is normatively very unappealing when that prerogative is posited to be independent of agent-centered restrictions. In any event, Scheffler needs to tell a much longer story than he has told in order to make the independence thesis a palatable ingredient in moral theory.

LARRY A. ALEXANDER

School of Law, University of San Diego

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

A Matter of Principle. RONALD DWORKIN. Cambridge, Mass.: Harvard, 1985. 425 p. $25.

Law's Empire. RONALD DWORKIN. Cambridge, Mass.: Harvard, 1986. 470 p. $20.

Despite having lived in England most of the past two decades, Ronald Dworkin is America's leading legal philosopher. His writings have helped raise the philosophical sophistication of academic lawyers and have demonstrated critical interrelations between legal philosophy and moral and political philosophy. His basic account of the nature of law and judicial decision, found in articles collected in *Taking Rights Seriously,*[1] competes for dominance with the explanation offered earlier in H. L. A. Hart's *Concept of Law.*[2] These two new volumes by Dworkin bear testimony to his eminence, evidencing his analytical ingenuity, powerful imagination, and elegant conceptualization. No subject ever seems quite the same after one has read Dworkin's treatment of it.

A Matter of Principle includes nineteen essays, written since *Taking Rights Seriously,* which cover a variety of topics and are written at different levels of difficulty. Seven essays, published originally for the *New York Review of Books,* address problems of public significance, particularly preferential treatment for minorities and press claims for First Amendment protection. These essays are, at least in their main features, comfortably comprehensible to a general readership. Some of the other essays require more work, especially if the reader is not closely familiar with previous writings of Dworkin and the people he attacks. The book's first part, which contains "A Forum of Principle," one of the most important pieces ever written about American constitutional adjudication, argues that judges deciding cases in both Britain and the United States inevitably make political decisions. This part also contains an illuminating analysis of justifications for civil disobedience. In the second part, Dworkin makes a technical defense of his familiar thesis that even controversial legal cases have one correct answer and develops analogies between literary interpretation and legal interpretation. The third part claims that liberalism is grounded on the notions that government should treat its citizens as equal, should be neutral among conceptions of the good, and should disregard external preferences, i.e.,

[1] Cambridge, Mass.: Harvard, 1977. [2] New York: Oxford, 1961.

0022-362X/87/8405/0284 $00.80 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

preferences that people have about what others do or have. Though the idea that government should be neutral among conceptions of the good might seem inhospitable to government support of high culture, Dworkin finds a justification for public support of the arts in the idea of preserving diverse possibilities for future generations. Part Four contains a devastating attack on the idea of wealth as a value and makes a complex argument that equality is the soundest basis for the belief that courts sometimes should decide on grounds of economic efficiency. In addition to defending preferential treatment as consonant with proper ideas of equality, the fifth part of the book has a fascinating exegesis of the problems surrounding the idea of legislative intent. The sixth part discusses First Amendment cases and Britain's 1979 Report of the Committee on Obscenity and Film Censorship, arguing in response to the latter that pornography should be dealt with in light of claims of justice rather than any "goal based" strategy.

When essays are written over a number of years and for different audiences, some strains are inevitable, and this book is not free of them. I shall mention two examples. Dworkin sharply attacks the dissent of now Chief Justice William Rehnquist in *United Steelworkers* v. *Weber,* which argued that remarks by sponsors showed that Congress's ban on racial discrimination in employment was meant to preclude voluntary programs that increase black employment by giving a set percentage of positions to blacks. An important part of Dworkin's attack on Rehnquist is that the *hopes* of most of those who voted for the act cannot be gleaned from the sponsors' statements. A preceding, though later written, chapter of the book contains a more complex analysis of bases for determining legislative intent and shows that a full challenge to Rehnquist's position requires appraisal of other possibilities.

The second example concerns Dworkin's argument that a newspaper reporter's claim to withhold sources of information should give way to a criminal defendant's need for the information. The reporter's claim, Dworkin says, is based on the *policy* of increasing information, whereas the defendant's claim is one of individual right, of *principle.* Courts should, Dworkin asserts, decide on grounds of principle, not policy. When Dworkin initially advanced this idea of the priority of principle over policy, critics responded that courts often, and properly, take account of broad consequential arguments. Dworkin replied that he did not mean to preclude all arguments of this sort; if the law authorized such inquiries as an aspect of determining an individual's rights, consequence-based arguments could be arguments of principle. The newspaper reporter, then, could put

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

his basic argument in the following way: "The First Amendment is a piece of constitutional legislation designed partly to serve the public policy of preserving the flow of information; it creates legal rights in those whose activities have this effect." So put, the argument in favor of withholding sources is turned into one of principle. That argument may be disposed of by a careful analysis of the scope of the First Amendment, but it cannot be rejected out of hand on the basis that it is self-evidently an argument of policy, not principle. The fuller First Amendment analysis would, I believe, support Dworkin's final position, and perhaps he implicitly assumes such an analysis; but this penultimate chapter of the book presently gives the impression that in practical cases Dworkin seeks more, and easier, mileage out of the principles/policies distinction than his more cautiously stated theoretical formulations of the distinction warrant.

Law's Empire, new rather than reprinted work, and indispensable reading for anyone seeking to understand modern jurisprudence, contains a systematic elaboration of Dworkin's views on law and legal interpretation. In its major features, his account maintains continuity with "Hard Cases," a 1975 essay which is a chapter of *Taking Rights Seriously,* but many points are clearer or more fully developed here. There is, for example, a complete theory of statutory interpretation, barely sketched in the initial essay. In ways that will often not be evident to the uninitiated reader, many of the new formulations in the book make substantial efforts to meet challenges to Dworkin's theory raised over the past decade.

The book anchors Dworkin's theory of law and legal interpretation in a more general view of what it means to interpret social practices. His main points of reference here are theories of literary interpretation. (He does not explore hermeneutic theories concerning religiously authoritative texts, whose explicitly normative character resembles law more closely than do typical pieces of literature.) Dworkin contends that any attempt to say what a work of literature or art means implicitly employs a view of what is good art. Similarly, interpretation of the meaning of legal norms requires political decision about the purposes of law. Arguing, against competing models, that law within countries like the United States and Great Britain is simply a matter of what has been conventionally settled (conventionalism) or that legal decision should be governed by what will work best in the future (pragmatism), Dworkin proposes an interpretation of law as *integrity.* Integrity is trying "to make the total set of laws morally coherent" (176).

This principle of integrity is a political virtue for legislatures, and it governs adjudication. In the latter context it requires courts to test

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

any proposed legal interpretation to see how it fits with other rules and principles of law and to choose, among those which fit adequately, the interpretation that reflects the soundest political morality. Dworkin's claim is that this notion of integrity both conforms with what courts try to do and provides a sound normative guide for their efforts. Although the gist of this position is found in "Hard Cases," the explication here is much fuller.

Building from his theory of statutory interpretation, Dworkin challenges the long-dominant assumption in constitutional scholarship of a critical division between interpretivist (or originalist) and noninterpretive (or nonoriginalist) approaches. He argues that all approaches are interpretive of what the Constitution says and all require some independent political judgments. For each area of legal interpretation—common law, statutory, and constitutional—Dworkin analyzes an actual case and runs though the analysis that Hercules, an ideal judge, would undertake. These exercises help one to understand the practical import of Dworkin's theory.

This parsimonious summary of the book's contents not only neglects the richness of many parts of the analysis, it omits altogether some remarkable discussions, such as Dworkin's ideas of "true community" and what makes a state legitimate. I have chosen in this short review to concentrate on the central question of Dworkin's theory of law and adjudication and how it now stands against its major alternative.

Dworkin first emerged as a critic of two fundamental claims of Hart's: that law in a country can be identified according to an ultimate rule of recognition implicitly accepted and applied by officials; and that, in situations in which existing legal criteria do not resolve a case, judges have a legislative-like discretion to make new law. Dworkin's response was that judges, contrary to then prevailing assumptions, are always trying to discover the law, not make it, and that what is law must be understood in terms of the set of principles that would best justify existing legal materials.

At least on one conceivable reading of "Hard Cases," Dworkin's theory seemed to have implications for what a judge should do rather radically different from those of Hart's theory. On this reading (clearly *dis*claimed by Dworkin within a few years after the appearance of "Hard Cases"), one set of political principles would fit the legal materials best, and the judge's legal duty would be to discover and apply that set of principles in deciding cases. In that event, decisions of difficult cases would require political judgments, but the content of these judgments could be drawn from the legal materials themselves. Decisions would not require *independent* judgments by

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Hercules, the ideal judge, about what is morally and politically best. Whereas the traditional idea of the judge as "legislating in the interstices" conjures up the notion of the judge deciding independently what would be right or desirable, or deferring to community opinion about those matters, Dworkin's model held the judge to the law itself. However difficult judges might find it to purge personal and community sentiment, that was what they should try to do. On this understanding, "law as discovery" really was at variance with "judicial legislation." Also on this understanding, a claim that virtually all legal problems have one correct answer seemed more modest than a claim that all moral problems or all choices about competing utilities have one correct answer, because the claim about law was grounded on the (often complicated) implications of concededly authoritative materials, sources not available for straightforward choices of morality and welfare.

Dworkin's approach in "Hard Cases," and even more so the amplified version in *Law's Empire,* illumines how judges seek to be guided by the law even when the materials most directly relevant, a particular statutory formulation or rule of common law, for example, seem inconclusive. Dworkin captures the reality that, most of the time, even in difficult cases, judges do not conceive of the law as "running out" and permitting them to undertake the separate task of legislation. (Some academic writing of judges suggests, however, that just such a perception occurs more often than Dworkin supposes.) No one can plausibly think that the law relevant to interpretation is exhausted when the materials most directly on point prove inconclusive. But someone might still suppose that Hart's theory is basically correct: that, in some cases (perhaps hard to identify), *all* the legal materials will be inconclusive, and then judges, whether they realize it or not, will be making new law. How different is Dworkin's present account from this theory?

During the past decade Dworkin has put decreasing emphasis on the work that "fit" will do in interpretation and increasing emphasis on the independent political and moral judgments of the judge, which determine which of the interpretations that fit adequately should be chosen. Perhaps his attention in *Law's Empire* to literary interpretation and to interpretation of a hypothetical social practice of courtesy underestimate the degree to which judgments about political morality can be drawn "neutrally" from legal materials, which, unlike most literature and customary practices of courtesy, themselves contain explicit normative criteria. In any event, a major theme of the book is that political decisions by judges, decisions not wholly determined by the legal materials themselves, are frequently

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

critical to judgment. Further, not only should judges themselves sometimes be deciding which is the more just of two competing interpretations of law, they should, in deference to the political virtue of "fairness," according to which community participation in lawmaking is valued, sometimes give weight to community judgments of what is just and desirable.

Two points are immediately evident. First, the thesis that virtually all legal cases have a right answer now depends on the assumption that virtually all moral questions have a right answer. Dworkin defends this position by claiming that, for people engaged in the interpretive enterprise, "external" skepticism about the status of moral judgments is irrelevant and by supposing that, within the enterprise, people proceed as if one answer really is better than another. (On earlier occasions, Dworkin has expressed skepticism, which I think counts as "internal," about whether interpersonal calculations of utility are possible. Since, in some situations, judges are supposed to decide which interpretation will best serve policies that lie behind legislation and since one function of legislation is to resolve competing claims of utility, I am not clear how judges can systematically avoid making such calculations, or whether Dworkin continues to think they should.)

The second point is that the standards that Dworkin thinks judges rely on to "interpret" look very much like the standards other theorists suppose judges use to "legislate in the interstices." This point leads to the question whether, from the judge's perspective, it now matters which of the two theories he or she accepts.

Any possibility that independent political principles should be brought to bear only at some grand theory-building stage seems implicitly rejected by *Law's Empire;* in their relevance to individual decisions the judge's political views will be as influential under Dworkin's theory as under Hart's. Perhaps the "fit" element of Dworkin's theory will eliminate some points of view that might otherwise be allowed. But if the view that all goods should be distributed equally is ineligible on "fit" grounds for American interpretation, on a genuinely disputed point of law the more relevant notion that goods should be distributed marginally more equally will almost certainly meet the demands of fit. Dworkin does contend that a judge cannot properly rely on extraneous grounds to resolve a case; he cannot use a statute designed to protect rare species to stop a dam he thinks unfortunate on aesthetic grounds. But because of the generality of rules and principles of law (any rule drawn in terms of rare species would have implications for projects not aesthetically undesirable), it seems unlikely that anyone recommending judicial legisla-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tion would approve of reliance on such extraneous grounds. Once one acknowledges that judges will not usually perceive the law as "running out," that in their considerations the legal materials, justice, and social desirability will be intertwined, it is not so clear whether the issue of judicial legislation or discovery retains much practical significance.

One important practical claim is at least implicit in Dworkin's basic account of interpretation as integrity. In rejecting the alternative of pragmatism, Dworkin effectively proposes that it can never be legally correct for a judge to promote desirable future consequences at the expense of the decision indicated by integrity. There are two serious weaknesses in Dworkin's defense of this position. The first is in the way he sets up the contest between the competing points of view. Let us suppose, as I do, that the law-as-integrity model is more appealing over all than the idea that, for a judge, everything comes down to making the best possible decision for the future. Why must the approaches be wholly and mutually exclusive? Why cannot one say that law as integrity can occasionally be qualified if the consequences would be bad enough? Many people assume that the best possible personal morality includes deontological and consequentialist elements which interrelate in ways not precisely definable; Dworkin really does not come to terms with the possibility of a similarily pluralist account of legal interpretation.

The second weakness is perhaps the book's greatest fault: the failure to integrate what is said with an important component of an appellate judge's role in the United States, the search for majority opinions. Dworkin's general position about conscious deviations from the result required by law as integrity is that such deviations are unambiguously *contrary to law* (even if occasionally morally justified). The treatment of his illustrative cases shows that the level of legal judgment in which he is primarily interested is neither the result nor the broad reasons for the result, but rather a statement of the applicable legal rule. In trying to achieve a majority, American appellate judges frequently settle for a statement of the applicable rule different from the formulation they would arrive at if they were sitting alone. Dworkin indicates approval of this practice as entirely appropriate for judges, but says it has to do with something other than deciding what the law as integrity requires. Perhaps some expansion of legal interpretation as integrity could encompass this practice of compromise; but in the absence of some such connection, we are left with two troubling questions: Why, according to Dworkin's theory, is a conscious compromise to get a majority opinion not

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

as *contrary to law* as any other conscious choice not to vote for what law as integrity requires? Why is pragmatic concern for effective opinions an aspect of judicial duty which makes deviation permissible when other pragmatic concerns do not warrant deviation?

Even if, from the inside—the position of the person interpreting the law—Dworkin's present account of what the judge may do does not seem sharply different in its practical implications from ideas of judicial "legislation," that does not detract from its appeal. Indeed, its ability to conceptualize how judges, scholars, and students tend to approach actual unresolved legal problems is one of its greatest strengths.

From the "outside," the question of legislation versus discovery looks somewhat different. I shall briefly address two dimensions. Suppose someone strongly objects to a particular judicial decision. To tell him that the preexisting law *required* the result he dislikes is somewhat misleading, if what one means is that the critical determinant is a sound political judgment that "fits" no better with the legal materials than the objector's contrary political judgment. Perhaps as a shorthand explanation to ordinary citizens, neither the notion that judges discover law nor the notion that they make law serves very well for difficult cases. There may be no substitute for a fuller explanation of the interpretive enterprise.

What of the sociologist trying to say what law is in a particular society? Dworkin rejects a conventionalist account of law, which he rightly says is no help for difficult problems of interpretation, since these reach issues that convention does not settle; but Dworkin admits that many premises about law, e.g., the supremacy of federal law over state law, are not, in any practical way, open to dispute. Though his own focus on the internal perspective is fully appropriate, what he says does not undermine the idea that the sociologist, or outsider, could manage pretty well with something like a conventionalist account.

A measure of the importance of *Law's Empire* and of the complexity and diversity of the ideas it explores is that these truncated critical comments leave altogether untouched many other fruitful sources for discussion.

KENT GREENAWALT

Columbia University, School of Law

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

Studies in Moral Philosophy is a newly established book series to be published by Peter Lang Publishing, under the editorship of John Kekes. The editor invites proposals for original book-length manuscripts whose primary focus is on the nature of good lives, character and its development, virtues and vices; the consideration of these and related topics in the light of literary works is particularly welcome. Manuscripts should be written in plain English, readily accessible to nonspecialists. Proposals, inquiries, and other correspondence should be sent to John Kekes, Department of Philosophy, SUNY at Albany, N.Y. 12222.

The Philosophy Department of Occidental College is very happy to announce the appointment of Saul Traiger (Pittsburgh) as Assistant Professor of Philosophy and Co-Director of the new interdisciplinary program in Cognitive Science.

The editors are pleased to announce a conference on Biomedical Ethics: An Anglo-American Dialogue, to be held June 22–24, 1987 at the Omni Park Central, New York City. The conference is jointly sponsored by the New York Academy of Sciences, the Hastings Center, the Royal Society of Medicine, and the Royal Society of Medicine Foundation. Experts in Biomedical Ethics from the United States and the United Kingdom will examine several significant ethical challenges confronting the medical communities and societies of both nations. Topics to be addressed include: the place and role of medicine in each culture; the balance of ethics, law, and medicine; the relationship between physician and patient; the challenges of new reproductive technologies; treatment decisions regarding newborns; allocation of scarce health-care resources; ethical issues in human-subjects research. Further information may be obtained from: The New York Academy of Sciences, Conference Dept., 2 East 63rd Street, New York, New York 10021.

The Society for Health and Human Values announces a call for papers for its 1987 Annual Meeting, to be held November 6–8, 1987, at the Arlington Hilton Hotel, Arlington, Va. The Society welcomes submissions on topics in the humanities and health sciences, including but not restricted to philosophy, ethics, history, literature, creative arts, and law. More information may be obtained from Andrew Jameton, Annual Meeting Program chair, University of Nebraska Medical Center, Department of Medical Jurisprudence and Humanities, 42nd and Dewey Avenue, Omaha, Neb. 68105-1065.

The editors report with sorrow the untimely death of Merrill Bristow Hintikka. Born in 1939, Professor Hintikka received a B.A. from Mills College in 1961 and a Ph.D. from Stanford University in 1982. She published numerous books and articles, and held a post of Associate Professor of Philosophy at Florida State University from December 1978 to the time of her death.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIV, NO. 6, JUNE 1987

IDENTITY, ESSENCE, AND INDISCERNIBILITY*

CAN things be identical *as a matter of fact* without being *necessarily* identical? Until recently it seemed they could, but now "the dark doctrine of a relation of 'contingent identity' "[1] has fallen into disrepute. In fact, the doctrine is worse than disreputable. By most current reckonings, it is refutable. That is, philosophers have *discovered* that things can never be contingently identical. Appearances to the contrary, once thought plentiful and decisive, are blamed on the befuddling influence of a powerful alliance of philosophical errors. How has this come about? Most of the credit goes to a simple argument (original with Ruth Marcus, but revived by Saul Kripke) purporting to *show* that things can never be only contingently identical. Suppose that α and β are identical. Then they share all their properties. Since one of β's properties is that necessarily it is identical with β, this must be one of α's properties too. So necessarily α is identical with β, and it follows that α and β cannot have been only contingently identical.[2]

* Donald Davidson, Sally Haslanger, Kit Fine, David Kaplan, Noa Latham, Shaughan Lavine, Barry Loewer, George Myro, Sydney Shoemaker, Robert Stalnaker, and David Velleman all made comments that helped me with the writing of this paper.

[1] Saul Kripke, *Naming and Necessity* (Cambridge, Mass.: Harvard, 1980), p. 4.

[2] In the prevailing necessitarian euphoria, it has become difficult to recapture the atmosphere of a few years back, when contingent identity was a commonplace of logical and metaphysical theorizing. To cite just two examples, Dana Scott's "Advice on Modal Logic" [in Karel Lambert, ed., *Philosophical Problems in Logic* (Boston: D. Reidel, 1970)] urged that "two individuals that are generally distinct might share all the same properties (of a certain kind!) with respect to the present world . . . Hence they are equivalent or *incident* at the moment. Relative to other points of reference they may cease to be incident" (165). And most of the early mind/body-identity theorists—U. T. Place, J. J. C. Smart, Thomas Nagel, among others—took themselves to be asserting the contingent identity of mental and physical entities. Smart, for instance, says very explicitly that "on the brain-process thesis the identity between the brain process and the experience is a contingent one" ["Sensations and

0022-362X/87/8406/0293$02.20
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar
© 1987 The Journal of Philosophy, Inc.

Digitized by Arya Samaj Foundation Chennai and eGangotri

I. A PARADOX OF ESSENTIALISM

Despite the argument's simplicity and apparent cogency, somehow, as Kripke observes, "its conclusion . . . has often been regarded as highly paradoxical."[3] No doubt there are a number of bad reasons for this (Kripke himself has exposed several), but there is also a good one: essentialism without some form of contingent identity is an untenable doctrine, because essentialism has a shortcoming that only some form of contingent identity can rectify. The purpose of this paper is to explain, first, why contingent identity is required by essentialism and, second, how contingent identity is permitted by essentialism.

Essentialism's problem is simple. Identicals are indiscernible, and so discernibles are distinct. Thus, if α has a property necessarily which β has only accidentally, then α is distinct from β. In the usual example, there is a bust of Aristotle, and it is formed of a certain hunk of wax. (Assume for the sake of argument that the hunk of wax composes the bust throughout their common duration, so that temporal differences are not in question.) If the bust of Aristotle is

---

Brain Processes," *Philosophical Review*, LXVIII, 2 (April 1959):141–156, p. 152)]. There is a question, actually, how it is that so many people *thought* that an impossible thing was possible. One hypothesis—maybe it is Kripke's hypothesis—is that these people were just very mixed up. And, in fact, it does seem that to varying degrees they were. Ruth Marcus once described W. V. Quine as thinking that modal logic was conceived in sin, the sin of confusing use and mention. Contingent identity had, if anything, even shadier beginnings, because at least three separate sins attended at its conception. Contingency was routinely identified with (or at least thought to follow from) *a posteriority;* particular identity statements (like "this pain is identical to that brain-event") were insufficiently distinguished from general identity statements (like "consciousness is a process in the brain"); and the contingent truth of an identity statement was equated with the contingency of the asserted identity, guaranteeing that contingent coincidence of concepts would be taken for the contingent identity of the things specified. So evidence for the confusion hypothesis is not lacking. The other hypothesis is, of course, that people recognized, confusedly perhaps, something sensible and defensible in the notion that things can be identical as a matter of fact. Philosophically, it does not much matter which of these hypotheses is correct. There *is* something sensible and defensible in the idea of contingent identity, whether its advocates recognized it or not. Or so I hope to show.

[3] Kripke, "Identity and Necessity," in Stephen P. Schwartz, ed., *Naming, Necessity, and Natural Kinds* (Ithaca, N.Y.: Cornell, 1977), p. 67. Let me say at the outset that, as far as I can see, Kripke's argument does succeed in establishing what it claims to establish, namely that identity, in the strict sense, can never obtain contingently. If the conclusion seems paradoxical, as it surely does, that is because people are confusing it with the genuinely paradoxical thesis that there can be *no* relation with the characteristics traditionally associated with "contingent identity." Speaking more generally, what Kripke says about identity is important and correct, and not questioned here; but people may have thought that his conclusions closed off certain avenues of investigation which are in fact still open. And that is perhaps why some of those conclusions have seemed hard to accept.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

necessarily a bust of Aristotle and if the hunk of wax is only *accidentally* a bust of Aristotle, then the bust and the hunk of wax are not the same thing. Or suppose that Jones drives home at high speed. Assuming that her speeding home is something *essentially* done at high speed, whereas her driving home only *happens* to be done at high speed, her speeding home and her driving home are distinct.

So far, so good, maybe; but it would be incredible to call the bust and the wax, or the driving home and the speeding home, distinct, and leave the matter there. In the first place, that would be to leave relations between the bust and the hunk of wax on a par with either's relations to the common run of *other* things, for example, the Treaty of Versailles. Secondly, so far it seems an extraordinarily baffling metaphysical coincidence that bust and wax, though entirely distinct, nevertheless manage to be *exactly alike* in almost every ordinary respect: size, weight, color, shape, location, smell, taste, and so on indefinitely. If distinct *statues* (say) were as similar as this, we would be shocked and amazed, not to say incredulous. How is such a coincidence possible? And, thirdly, if the bust and the wax are distinct (pure and simple), how is it that the number of middle-sized objects on the marble base is not (purely and simply) 2 (or more)? Ultimately, though, none of these arguments is really needed: that the bust and the wax are in *some* sense the same thing is perfectly obvious.

Thus, if essentialism is to be at all plausible, nonidentity had better be compatible with intimate identity-*like* connections. But these connections threaten to be inexplicable on essentialist principles, and essentialists have so far done nothing to address the threat.[4] Not

[4] Observing that not only modal but temporal differences "establish that a statue is not the hunk of stone, or the congery of molecules, of which it is composed," Kripke allows that "mere non-identity . . . may be a weak conclusion" ("Identity and Necessity," p. 101). *Extremely* weak, from the point of view of philosophical materialism. That pains were not identical with neural stimulations seemed to be a powerfully antimaterialistic result; but now it turns out to be compatible with pains and neural stimulations' being as tightly bound up with one another as statues and their clay. And what materialist would not be delighted with that result? On the other hand, "The Cartesian modal argument . . . surely can be deployed to maintain relevant stronger conclusions as well" ("Identity and Necessity," p. 101). Possibly this means that the statue is "nothing over and above" its matter, whereas the same cannot be said of a person; in the sense that necessarily, the statue (but not the person) exists if its matter does, and with a certain organization (*Naming and Necessity*, p. 145). But it seems doubtful whether the statue *is* "nothing over and above" its matter *in that sense* (what if the statue's matter had gathered together by chance, before the earth was formed? what if a different sculptor had organized the matter?); and the subtler the sense in which a statue really is "nothing over and above" its matter, the less implausible it becomes that, in a substantially similar sense, a person is "nothing over and above" *its* matter. So there may still be room for doubt whether modal arguments establish significantly more difference between a person and her matter than between a statue and its.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

quite nothing, actually; for essentialists have tried to understand certain (special) of these connections in a number of (special) ways. Thus, it has been proposed that the hunk of wax *composes* the bust; that the driving home *generates* the speeding home; that a neural event *subserves* the corresponding pain; that a computer's structural state *instantiates* its computational state; that humankind *comprises* personkind; and that a society is *nothing over and above* its members. Now all these are important relations, and each is importantly different from the others. But it is impossible to ignore the fact that they seem to reflect something quite general, something not adequately illuminated by the enumeration of its special cases, namely, the phenomenon of things' being distinct *by nature* but the same *in the circumstances*. And what is that if not the—arguedly impossible—phenomenon of things' being contingently identical but not necessarily so? The point is that, if essentialism is true, then many things that are obviously in *some* sense the same will emerge as strictly distinct; so essentialism must at least provide for the possibility of intimate identity-*like* connections between distinct things; and such connections seem to be ways of being contingently identical. Essentialism, if it is to be plausible, has to be tempered by some variety of contingent identity.

Hence, essentialism is confronted with a kind of paradox: to be believable it needs contingent identity; yet its principles appear to entail that contingent identity is not possible. To resolve the paradox, we have to ask: What is the "nature" of a particular thing?

II. ESSENCE

Begin with a particular thing α. How should α be characterized? That is, what style of characterization would best bring out "what α is"? Presumably a characterization of any sort will be via certain of α's properties. But which ones?

Why not begin with the set of all α's properties whatsoever, or what may be called the *complete profile* of α? Since α's properties include, among others, that of being identical with α, there can be no question about the sufficiency of characterization by complete profile. But there may be doubt about its philosophical interest. For the properties of α will generally be of two kinds: those which α had to have and those which it merely happens to have. And, intuitively, the properties α merely happens to have reveal nothing of what α is, as contrasted with what it happens to be like. As Antoine Arnauld explains in a letter to Leibniz,

> . . . it seems to me that I must consider as contained in the individual concept of myself only that which is such that I should no longer be me if

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

> it were not in me: and that all that is to the contrary such that it could be or not be in me without my ceasing to be me, cannot be considered as being contained in my individual concept.[5]

(Adding: "That is my idea, which I think conforms to everything which has ever been believed by all the philosophers in the world"!) If α's nonnecessary properties reveal nothing about what α is, nothing will be lost if they are struck from its characterization.

Dropping α's nonnecessary properties from its complete profile yields the set of all properties that α possesses essentially, or what can be called the *complete essence* of α.[6] Since α is essentially identical with α, the property of so being will be included in α's complete essence; so the sufficiency of the characterization is again beyond doubt. Nor can there be much question that complete essences do better than complete profiles at showing what particulars are by nature. But worries about philosophical interest remain.

In the first place, the essence of an entity ought, one feels, to be an assortment of properties *in virtue of which* it is the entity in question. But this requirement is trivialized by the inclusion, in essences, of identity properties, like that of being identical with California. A thing does not get to be identical with California by having the property of so being, but gets to be California and to have that property, alike, by having certain *other* properties. And it is these other properties that really belong in a thing's characterization. Another way of putting what is probably the same point is that identity properties and their ilk are not "ground floor," but dependent or supervenient. As a kind of joke, someone I know explains the difference between his two twin collies like this: "It's simple: *this* one's *Lassie*, and *that* one's *Scottie*." What makes this a joke is that that cannot be all there is to it; and the reason is that identity properties are possessed not *simpliciter*, but dependently on other properties. It is only these latter properties that ought, really, to be employed in a thing's characterization.

Secondly, the essence of a thing is supposed to be a measure of *what is required* for it to be that thing. But, intuitively, requirements can be more or less. If the requirements for being β are stricter than the requirements for being α, then β ought to have a "bigger" essence than α. To be the Shroud of Turin, for instance, a thing has to

[5] H. T. Mason, ed., *The Leibniz-Arnauld Correspondence* (Manchester: University Press, 1967), p. 30.

[6] In some contexts it is useful to distinguish between essential and necessary and between accidental and contingent properties. For example, someone might think that, whereas Socrates is essentially human, he is only necessarily Greek-or-not. The distinction is intuitive but irrelevant to our purposes.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
have everything it takes to be the associated piece of cloth, *and* it has to have enshrouded Jesus Christ (this is assuming that the Cloth of Turin did, in fact, enshroud Jesus Christ). Thus, more is essential to the Shroud of Turin than to the piece of cloth, and the Shroud of Turin ought accordingly to have the bigger essence. So, if essences are to set out the requirements for being their possessors, it should be possible for one thing's essence to include another's.[7] What is perhaps surprising, however, is that this natural perspective on things will not survive the introduction of identity properties and their ilk into individual essences. Think of the piece of cloth that makes up the Shroud of Turin (call it "the Cloth of Turin"): if the property of being identical with the Cloth of Turin is in the Cloth of Turin's essence, then, since that property is certainly not in the *Shroud* of Turin's essence, the inclusion is lost. Equivalently, it ought to be possible to start with the essence of the Cloth of Turin, *add* the property of having served as the burial shroud of Jesus Christ (along perhaps with others this entails), and wind up with the essence of the Shroud of Turin. But, if the property of being identical to the Cloth of Turin is allowed into the Cloth of Turin's essence, then adding the property of having served as Jesus's burial shroud produces a sort of contradiction; for, obviously, *nothing* is both identical to the Cloth of Turin and necessarily possessed of a property—having served as Jesus's burial shroud—which the Cloth of Turin possesses only contingently. And the argument is perfectly general: if identity proper-

[7] Where *kinds* of things are concerned, this is comparatively uncontroversial: the essence attaching to the kind *cow* strictly includes the essence attaching to the kind *animal*. But, as Leibniz noticed, individuals can be thought of as instancing smallest or least kinds, what we might call *individual kinds* (what sets *individual* kinds apart is that in each possible world at most one thing instances them): ". . . since St. Thomas could maintain that every separate intelligence differed in kind from every other, what evil will there be in saying the same of every person and in conceiving individuals as final species" [G. R. Montgomery, ed., *Discourse on Metaphysics, Correspondence with Arnauld, Monadology* (La Salle, Ill.: Open Court, 1908), p. 237]. Now just as the essences of general kinds can be comparable, so can the essences of an individual and a general kind (*Bossie*'s essence includes that of *cow*). But then why should the essences of individual kinds not be comparable too? There is every reason to see the relation between the Shroud of Turin and the piece of cloth as continuous with that between cow and animal: just as it is harder to be a cow than an animal, it is harder to be the Shroud of Turin than the piece of cloth, and just as nothing can be a cow without being an animal (but not conversely), nothing can be the Shroud of Turin without being the piece of cloth (but not conversely). So there seems to be a strong case for extending the familiar doctrine that the essence of one kind can include that of another to individual kinds, and, what comes to the same, to individuals themselves. (Incidentally, I am assuming that the Shroud of Turin could not have been made of anything other than that piece of cloth. Something made of another piece of cloth might have been called "the Shroud of Turin," but it would not have been our Shroud of Turin.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ties (or others like them) are allowed into things' essences, then distinct things' essences will always be incomparable.[8]

Implicit in the foregoing is a distinction between two types of property. On the one hand, there are properties that can only "build up" the essences in which they figure. Since to include such properties in an essence is not (except trivially) to keep any other property out, they will be called *cumulative*. On the other hand, there are properties that exercise an inhibiting effect on the essences to which they belong. To include this sort of property in an essence is always to block the entry of certain of its colleagues. Properties like these —identity properties, kind properties, and others—are *restrictive*. If restrictive properties are barred from essences, that will ensure that essences are comparable, and so preserve the intuition that each essence specifies what it takes to be the thing that has it.

Essences constrained to include only cumulative properties will have two advantages. First, they will determine their possessors' inessential properties negatively, not by what they include but by what they leave out; and, as a result, things' essences will be amenable to expansion into the larger essences of things it is "more difficult to be," thus preserving the intuition that a thing's essence specifies what it takes to be that thing. And, second, things will be the things they are *in virtue of* having the essences they have. To put it approximately but vividly, they will be what they are because of what they are like (see Prop. 4).[9] Our tactic will be to look first for properties suited

[8] Identity properties are by no means the only properties that lead to these difficulties. Kind properties, for example, are just as bad. If the property of being a piece of cloth (i.e., being of the *kind* piece of cloth) is included in the Cloth of Turin's essence, then adding on the property of having served as Jesus's burial shroud (along with perhaps some others) can no longer yield the essence of the Shroud of Turin. For it is never essential to any piece of cloth that it should have been used in any particular way (necessarily, any piece of cloth could have been destroyed moments after its fabrication). Incidentally, kind properties are disqualified by the first argument too: like identity properties, they are possessed not *simpliciter*, but dependently on other properties. It defies credulity that two things should be indiscernible up to this detail, that one is a collie and the other is not. (Thinking of identity and kind properties as *classificatory*, rather than *characterizing*, the above becomes the truism that a thing's classification depends entirely on what the thing is like.)

[9] Although this is probably overstating it, at least as far as what is actually established goes (see Prop. 4 below). See also David Wiggins, *Sameness and Substance* (Cambridge, Mass.: Harvard, 1980), and Robert Merrihew Adams, "Primitive Thisness and Primitive Identity," this JOURNAL, LXXVI, 1 (January 1979): 5–26, in both of which the sufficiency of "quality" for "quiddity" is considered and rejected. Wiggins's opinion is that

> . . . to make clear which thing a thing is, it is not enough (*pace* the friends of the logically particularized essence) to say however lengthily that it is *such*, or *so and so*. We have to say that it is *this* or *that such*. This is perfectly obvious when

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
to inclusion in cumulative essences and then to show that, under reasonable further assumptions, identity supervenes on cumulative essence.

III. MODELING ESSENCE

To find a set of properties suitable for the construction of cumulative essences, one needs to know what "properties" are; especially because a totally unrestricted notion of property is incoherent, as Richard's and Grelling's paradoxes show.[10] So it makes sense to look for a sharper formulation of the notion of property before pushing ahead with the search for cumulative essence. Such a formulation is provided by the apparatus of possible worlds.

Let $\mathcal{L}$ be an ordinary first-order language with identity, and let $\mathcal{L}(\Box)$ be $\mathcal{L}$ supplemented with the sentential necessity operator '$\Box$'. To a first approximation, a model of $\mathcal{L}(\Box)$ is just a set $\mathcal{W}$ of models W of $\mathcal{L}$ (to be thought of as possible worlds). But there is a qualification. Traditionally, a model's domain is simultaneously the set of things that *can be talked about* and the set of things that *exist*, i.e., the domain of *discourse* and the *ontological* domain. But, since one can talk about things that do not exist, W's domain of discourse should be allowed to contain things not in its ontological domain; and since there are not, mystical considerations to the side, things about which one cannot talk, W's ontological domain should be a subset of its domain of discourse. What this means formally is that with each model W in $\mathcal{W}$ is associated a subset $\mathcal{D}(W)$ of its domain (intuitively, the set of things existing in W). Let W thus supplemented be known as a *free model* of $\mathcal{L}$. For simplicity's sake, every member of $\mathcal{W}$ will have the same domain $\mathcal{D}$, and $\mathcal{D}$ will be the union of the $\mathcal{D}(W)$'s. And now a model of $\mathcal{L}(\Box)$ can be defined as a set $\mathcal{W}$ of free models of $\mathcal{L}$, such that the domain of discourse of each is the union of all their ontological domains.[11]

---

we think of trying to determine one entity by mentioning short or simple predicates (other than *identical with x* or such like). But it is difficult to see any reason to believe that by making ordinary predicates ever longer and more complicated we shall be able to overcome the obvious non-sufficiency or non-necessity for identity with just *x* that infects all the relatively simple predicates true of *x* (104/5).

[10] See Peter Geach, "Identity," in his *Logic Matters* (Oxford: Blackwell, 1972). Baruch Brody asserts that his theory, according to which items are identical if and only if they are indiscernible over "all" properties, "is not ruled out by its leading to any paradoxes" [*Identity and Essence* (Princeton, N.J.: University Press, 1980), p. 18]. But he does not satisfactorily explain why not.

[11] The accessibility relation is omitted; in effect, every world has access to every other.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Tempting though it is to define a property as any function P from worlds W to subsets $P(W)$ of $\mathcal{D}$, there is reason not to. For when will α have P necessarily: when it has P in every world, or when it has it in every world in which it exists? Not the former, because then everything necessarily exists.[12] Nor the latter, first, because it permits a thing to possess only accidentally a property it must perish to lose and, second, because it upsets the principle that essence varies inversely with existence, i.e., the fewer the worlds a thing exists in, the more properties it has essentially. What this in fact points up is a difference between two kinds of characteristic: being human in every world where you exist is sufficient for being human everywhere (almost all characteristics are like this), but existing in every world where you exist is obviously not sufficient for existing everywhere (apparently only existence and characteristics involving existence are like this). From now on, an *attribute* is a function from worlds W to subsets of $\mathcal{D}$, and a *property* is an attribute P such that anything having it wherever it exists has it everywhere. In general, an attribute is necessary to a thing if it attaches to the thing in every possible world (preserving the intuition that existence is sometimes contingent). If the attribute is also a property, this reduces to the thing's having the attribute wherever it exists (preserving the intuition that humanity is necessary to Socrates if he cannot exist without it). In what follows, properties (rather than attributes in general) are the items under investigation.

From the definition of property, it follows that, if P is a property, then so are $P^{\Box}{:}W \rightarrow \{\alpha\epsilon\mathcal{D}|\forall W'\alpha\epsilon P(W')\}$ (the property of being essentially P, or P's *essentialization*); $P^{\Diamond}{:}W \rightarrow \{\alpha\epsilon\mathcal{D}|\exists W'\alpha\epsilon P(W')\}$ (the property of being possibly P, or P's *possibilization*); and $P^{\triangle}{:}W \rightarrow \{\alpha\epsilon P(W)|\exists W'\ \alpha\epsilon P(W')\}$ (the property of being accidentally P, henceforth P's *accidentalization*). The essentialization $X^{\Box}$ (accidentalization $X^{\triangle}$) of a set X of properties is the set of its members' essentializations (accidentalizations). If α is in $P(W)$ and exists in W, then it is in $P[W]$ (note the square brackets). If for each P in X $\alpha\epsilon P(W)$, then $\alpha\epsilon X(W)$; if, in addition, α exists in W, then it is in

[12] The problem existence raises for the definition of 'essential' is not unfamiliar. Kripke alludes to it in "Identity and Necessity":

> Here is a lectern. A question which has often been raised in philosophy is: What are its essential properties? What properties . . . are such that this object has to have them if it exists at all . . . [Footnote:] This definition is the usual formulation of the notion of essential property, but an exception must be made for existence itself: on the definition given, existence would be trivially essential. We should regard existence as essential to an object only if the object necessarily exists. (86)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

$X[W]$. A set Y of properties is *satisfiable* in W, written $\mathrm{Sat}[Y,W]$, iff there is something in $\cap_{P \epsilon Y} P[W]$. Given a set X of properties, a thing α's *X-essence* $\mathbb{E}_X(\alpha)$ is the set of all P in X which α possesses essentially, or $\{P \epsilon X | \exists W(\alpha \epsilon P^{\Box}(W))\}$. β is an *X-refinement* of α, written $\alpha \leq \beta(X)$—or just $\alpha \leq \beta$ if X is clear from context—iff α's X-essence is a subset of β's, i.e., if $\mathbb{E}_X(\alpha) \subseteq \mathbb{E}_X(\beta)$.

That essences drawn from X should be amenable to expansion is a condition not on X alone, but on X and $\mathcal{W}$ taken together: X and $\mathcal{W}$ must be so related that suitably expanding the X-essence of any thing in any world in $\mathcal{W}$ always produces the X-essence of some other thing existing in that same world. Let $\Omega = \langle \mathcal{W}, X \rangle$ be a *property-model* of $\mathcal{L}(\Box)$ if $\mathcal{W}$ is a model of $\mathcal{L}(\Box)$ and X is a set of properties on $\mathcal{W}$. A property-model Ω is *upward-closed,* or *u-closed,* iff:

(U) $\forall \alpha\ \forall Y \subseteq X - \mathbb{E}_X(\alpha)\ \forall W\ [\alpha \epsilon Y^{\Delta}[W] \Rightarrow (\exists \beta \geq \alpha) \beta \epsilon Y^{\Box}[W]]$

In words, given any α, given any set Y of properties not essential to α, and given any world W, if α exists in W and has Y there, then it has a refinement β which exists in W and has Y essentially there. (For future reference, (U) is provably equivalent to the simpler statement that $\forall Z \subseteq X\ \forall W\ [\mathrm{Sat}[Z,W] \Rightarrow \mathrm{Sat}[Z^{\Box},W]]$.)

Upward closure requires that any existing α possessing (suitable) properties inessentially be refinable into an existing β that possesses those properties essentially. The converse is intuitive too: if there exists a β refining α which essentially possesses (suitable) properties not essential to α, then α should exist and possess those same properties accidentally. Thus, if the Shroud of Turin exists in a world W, then not only should the Cloth of Turin exist in W, but it should serve as Jesus's burial shroud in W. Not only is this plausible on the face of it, but otherwise it is hard to see what separates the worlds in which the Cloth occurs by itself from those in which it occurs together with the Shroud; whereas surely the difference is that in the latter, but not the former, the Cloth serves as Jesus's burial shroud. And, in general, if β refines α, then surely what separates worlds in which α exists without β from those in which α exists with β is that, in the latter worlds, α possesses the difference between their X-essences, whereas in the former it does not. Specifically, if β refines α, then (1) α exists wherever β does, (2) in worlds where both exist, α accidentally possesses every property in $\mathbb{E}_X(\beta) - \mathbb{E}_X(\alpha)$, i.e., every property in the difference between their essences, and (3) in worlds where just β exists, α does not possess all the properties in $\mathbb{E}_X(\beta) - \mathbb{E}_X(\alpha)$. All of these follow on the addition of a requirement of *downward closure,* literally the converse of the upward closure enforced above:

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(D) $\forall\alpha\ \forall Y \subseteq X - \mathbb{E}_X(\alpha)\ \forall W\ [(\exists\beta \geq \alpha)\beta\epsilon Y^{\Box}[W] \Rightarrow \alpha\epsilon Y^{\Delta}[W]]$

In words, for anything α, any properties Y not essential to it, and any world, if α has an existing refinement β possessing Y essentially, then α exists and possesses Y accidentally. Property-models satisfying both (U) and (D) are *closed*.

Prop. 1 Let Ω be closed. If $\alpha \leq \beta$, then

(1) $\mathcal{W}(\beta) \subseteq \mathcal{W}(\alpha)$
(2) $\forall W\epsilon\mathcal{W}(\beta)\ [\alpha\epsilon(\mathbb{E}_X(\beta) - \mathbb{E}_X(\alpha))^{\Delta}(W)]$
(3) $\forall W\epsilon\mathcal{W}(\alpha) - \mathcal{W}(\beta)\ [\alpha \notin (\mathbb{E}_X(\beta) - \mathbb{E}_X(\alpha))^{\Delta}(W)]$

Proof: For (1), observe first that $\Lambda[W] = \Lambda(W) \cap \mathcal{D}(W) = \mathcal{D} \cap \mathcal{D}(W) = \mathcal{D}(W)$ (because the null intersection is everything, in this case $\mathcal{D}$). By (D), $\beta\epsilon\mathcal{D}(W) \Rightarrow \beta\epsilon\Lambda[W] \Rightarrow \beta\epsilon\Lambda^{\Box}[W] \Rightarrow \alpha\epsilon\Lambda^{\Delta}[W] \Rightarrow \alpha\epsilon\Lambda[W] \Rightarrow \alpha\epsilon\mathcal{D}(W)$.

For (2), just let Z be $\mathbb{E}_X(\beta) - \mathbb{E}_X(\alpha)$. For (3), suppose that β does not exist in W, and suppose *per absurdum* that α accidentally possesses, in W, the difference between its X-essence and β's. By upward closure, α has a refinement γ which exists in W and which possesses the whole lot, i.e., all of $\mathbb{E}_X(\beta)$, essentially. But then γ refines β; so, by (1), β exists in W after all. Contradiction. □

IV. CONTINGENT IDENTITY

Things that disagree in any of their properties are not identical. The Shroud of Turin, which (let us suppose) *had* to enshroud Jesus, is thus distinct from the Cloth of Turin, which did not. But, as we said, there is something deeply troubling about leaving matters thus. After all, the Shroud of Turin is also distinct from the Treaty of Versailles. Do we really want to leave the Shroud's relations with the Cloth on the same level as its relations with the Treaty of Versailles? And the trouble does not stop here. The Cloth and the Shroud differ, it is true, but it must also be said that their differences are of a somewhat *recherché* variety. In every *ordinary* respect the two are exactly alike. And this is on the face of it a rather extraordinary coincidence. That the Cloth and the Shroud are specially connected seems undeniable, but something must be done to demystify the connection. If it is not identity, what is it?

Maybe the answer is that it *is* identity, but identity of a different, less demanding, character. In the terms of a currently unpopular theory—and notwithstanding the argument that seems to rule it out—it is "contingent identity," or (the more neutral term) "coincidence."[13] Despite the once widespread enthusiasm for contingent

[13] Another reason for preferring "coincident" to "contingently identical" is that properly identical things will also be coincident, indeed necessarily so, and it sounds funny to say that they are *necessarily* contingently identical. But I continue to use the term, partly for shock value, and partly for reasons to be given presently.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

identity, it seems to me that the idea never received a satisfactory formulation. Specifically, all the analyses I have seen have a drawback in common: they (sometimes explicitly, sometimes in effect) treat things as strung together out of their modal manifestations (states, slices, stages), and call them coincident in a world if their manifestations in that world are properly identical. There are two objections to this kind of explication. The first is that it relies, ultimately, on the notion to be explicated; for one has little idea what a thing's state or manifestation in a world is, if not something whose nature is exhausted by its being exactly like the thing so far as the relevant world is concerned, i.e., by its being contingently identical with the thing in that world. Even more important, intuitively, things (e.g., animals) are *not* strung together out of their modal manifestations in this way, and proper identity of modal manifestations is not what is meant by contingent identity. Intuitively, things are just, well, *things,* and coincidence is a matter of things' circumstantial sameness.

How, then, is circumstantial sameness to be separated out from total sameness? What marks off the "ordinary" respects in which the Cloth and the Shroud are alike from the "extraordinary" respects in which they differ? Let us start with Dana Scott's idea that "two individuals that are generally distinct might share all the same properties (of a certain kind!) with respect to the present world." ("Advice on Modal Logic," *op. cit.*, p. 165). Probably the most obvious way of elucidating this would be to say the following: α and β are contingently identical (in a world) if and only if they have the same *contingent properties* (in that world). So, for example, the bust and the hunk of wax agree in their size, weight, color, and so on—and all these are, of course, properties they have contingently. Maybe contingent identity is sameness of contingent properties.

That that cannot be right follows from the fact that, if anything has a property contingently, then it has all its stronger properties contingently too. (Suppose α has P contingently; then there is a world in which α lacks P; but if Q is stronger than P, α also lacks Q there; and, since α has Q, it has it contingently.) So, for example, if Paris is only contingently romantic, then it is only contingently identical-to-Paris-and-romantic. But that means that anything that has the same contingent properties that Paris has is (among other things) identical with Paris. And that already shows that Paris is the only thing with exactly its contingent properties. Thus, if contingent identity is treated as sameness of contingent properties, contingent identity collapses into identity proper.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Still, from a certain perspective, this first analysis might be only a little way off the mark. To the question, What makes a thing's possession of a property *circumstantial?* it seems natural to reply that the possession is circumstantial if it depends on *how matters actually stand with the thing*. But now notice that this is ambiguous. Depends *how: partly* or *wholly?* If you answer "partly," then you get the thing's *contingent* properties. But, if you answer "wholly," you get the properties the thing has *entirely* in virtue of how matters actually stand with it; and these properties, what can be called the *categorical* properties, seem intuitively to be the ones in question.[14] For if two things agree in all their categorical properties (in a world), then *so far as that world and it alone* is concerned, the two things are just the same. And that is what was meant by "contingent identity." So contingent identity is categorical indiscernibility.

To come to this conclusion from a different direction, consider again the driving home and the speeding home. What separates the "ordinary" respects in which these two are alike from the "extraordinary" respects in which they differ? For a start, the driving home could have been done slowly, but not so the speeding home; the driving home had higher prior probability than the speeding home; if the driving home had not occurred, Jones would have taken the bus home, but the same cannot be said of the speeding home; and the speeding home, rather than the driving home as such, caused Jones's accident. Thus, the driving home and the speeding home differ in—among other things—their modal, probabilistic, counterfactual, and causal properties. Now what is special about modal, probabilistic, counterfactual, and causal properties? Primarily this: they are grounded not just in how a thing *actually* is, but on how it *would* or *could* have been if circumstances had been different. All a thing's other properties, by contrast, are grounded entirely in how it is in the circumstances that happen to obtain. The former properties are a thing's *hypothetical* properties, the latter its *categorical* properties. Now the contingent identity of the driving home with the speeding home seems intuitively to be a matter of their sharing such properties as speed, place, time, etc., regardless of their modal, causal, probabilistic, and counterfactual differences; that is, their contingent identity seems to be a matter of their sharing their categorical properties, irrespective of their hypothetical differences.

[14] To be absolutely clear about the difference between contingency and categoricity, consider their complements. Where a thing has its noncontingent properties *necessarily*, it has its noncategorical properties *hypothetically*.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

How are a thing's categorical properties to be found? (Actually it will be simplest if we look for properties categorical *as such,* i.e., properties that can only be had in a manner independent of what would or could have happened.) Why not take the intuitive notion that a property is categorical just in case a thing's having it is independent of what goes on in nonactual worlds, and try to turn this into a definition? The problem is that such a definition would be circular. Suppose it is a categorical property of this hunk of clay that it is spherical. How can that depend on how the clay comports itself in other worlds? But, if you think about it, it does, in that the clay's being spherical in this world depends on its being, in those worlds, such that in *this* world it is spherical. The problem is that being such that, in this world, it is spherical, is a hypothetical property of the clay. So, apparently, what we really meant to say was that a property is categorical if it attaches to a thing regardless of its *categorical properties* in other worlds. And that is clearly circular.[15]

Somehow the circularity has to be circumvented. Things are going to be coincident in a world iff they have exactly the same categorical properties there. But maybe this can be turned around: the categorical properties are exactly those which cannot tell coincident things apart.

Postpone for a moment the question of how that would help, and ask, instead, is it even true? That is, *are* the categorical properties the properties insensitive to the difference between coincidents? This will be the case only if coincidence is compatible with every kind of hypothetical variation, i.e., if, for every hypothetical property, coincidents can be found that disagree on it. But it is clear that ordinary

[15] Two remarks. As a characterization of the categorical properties, the foregoing is circular. But it does have the virtue of illustrating why not *every* property can be hypothetical (as is sometimes suggested). For a property to be hypothetical, whether a thing has it must depend on the things' *categorical* properties in other worlds; and that shows that no property can be hypothetical unless at least some properties are categorical. And, since it is relatively unproblematic that categorical properties give rise to hypothetical properties (given the present broadly essentialist assumptions), neither category can be emptied without emptying the other. So a skeptic about the distinction should maintain that no property is of *either* kind, not that all (some) properties are of one kind and none are of the other. [For example, Sydney Shoemaker's theory of properties as "second-order powers," though it might seem to imply that all properties are hypothetical, or, on another reading, that all properties are categorical, is perhaps better read as rejecting the distinction altogether. See "Causality and Properties," and "Identity, Properties, and Causality," collected in his *Identity, Cause, and Mind* (New York: Cambridge, 1984).] Second, in rejecting the proposed account of categoricity on grounds of circularity, I do not mean to imply that the account I finally give is not itself ultimately circular. Given the cumulative properties, the categorical properties can be noncircularly specified; but the cumulative properties themselves cannot be noncircularly specified, in particular not by the formal conditions laid down above.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

things do not exhibit the hypothetical variety that this would require. Among ordinary things, coincidents never differ on (e.g.) the score of fragility (if the statue is fragile then so is the piece of clay); among ordinary things, one never finds one thing accidentally juvenile, or mature, and another, coincident with the first, essentially so (simply because *no* ordinary thing is essentially juvenile, or mature).

So much for ordinary things. But what about things as seen from the vantage point of metaphysics? Metaphysics aspires to understand reality as it is in itself, independently of the conceptual apparatus observers bring to bear on it. Even if we do not ourselves recognize essentially juvenile or mature entities, it is not hard to imagine others who would[16]; and to someone who, in addition to the statue and the piece of clay, discerned a statue-cum-shards, not everything coincident with the statue would be fragile. Conversely, we recognize things, say, essentially suitable for playing cribbage, or cutting grass, which others do not, or might not have. To insist on the credentials of the things *we* recognize against those which others do, or might, seems indefensibly parochial. In metaphysics, unusual hypothetical coloring can be no ground for exclusion.[17] Since this is metaphysics, everything up for recognition must actually be recognized; and, when this is done, there are coincidents enough to witness the hypotheticality of every hypothetical property.[18]

Given information about what was coincident with what, the categorical properties could be identified: they would be the properties insensitive to the difference between coincidents. Now, as of yet, there *is* no information about what is coincident with what (that is why we were looking for the categorical properties in the first place).

[16] To get a sense of what it might be like to countenance a creature coming into existence "in mid-life," consider Jane Eyre's reflections on the eve of her (anticipated) marriage to Mr. Rochester: "Mrs. Rochester! She did not exist: she would not be born till tomorrow, sometime after eight o'clock A.M.; and I would wait to be assured she had come into the world alive before I assigned her all that property." For a creature that stops existing "in mid-life," there are the opening lines of Neil Young's "A Child's Claim to Fame": "I am a child/I last a while."

[17] Less dogmatically, there are two kinds of metaphysics: descriptive and transcendental. In descriptive metaphysics one is interested in reality as people see it; in transcendental metaphysics one tries to abstract to the largest extent possible from the human contribution. Pretty clearly, the distinction is relative. All metaphysics is somewhat transcendental (metaphysicians do not spend much time thrashing out the nature of time zones), but probably the present approach is more transcendental than most.

[18] To say that everything up for recognition is actually recognized, is to say the following: given any set of worlds, and given an assignment to each of categorical properties satisfiable therein, there is something that exists in those worlds exactly and possesses in each the associated categorical properties. Call this the requirement of *fullness*.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

But that is not to say that none can be obtained; and, in fact, *certain* cases of coincidence—enough to weed out all the noncategoricals—are discoverable in advance.

To find these cases, try to imagine pairs of things that differ as little as possible from being strictly identical (for things almost identical will be contingently identical if any things are). Trivially, if α and β are strictly identical, α will exist in exactly the same possible worlds as β, and α will be coincident with β in all of them. To arrange for the least possible departure from this, let α exist in a few more worlds than β, but otherwise leave everything unchanged, i.e., let them be coincident in all the worlds where both exist. As it happens, that is exactly how it is with the driving home and the speeding home. Wherever the speeding home occurs, the driving home occurs too, and is coincident with the speeding home. But there will also be worlds in which the driving home is done at a reasonable speed, and in such worlds the speeding home does not occur.

Still, none of this helps with the project of explicating contingent identity, unless there is a way of characterizing the given relation—the relation between the driving home and the speeding home—which does not itself rely on the notion of contingent identity. But there is: it is the relation of refinement. Although only a fraction of all coincidents stand in the relation of refinement, this fraction is enough to weed out all the noncategorical properties.[19] With the noncategorical properties weeded out, the categorical properties are

[19] To see *why* the properties insensitive to the difference between things related by refinement can be relied on to be exactly the categoricals: Call these properties the *provisionally* categorical properties, and call things indiscernible with respect to these properties *provisionally* coincident. The problem is really to prove that every provisionally categorical property is genuinely categorical (the converse is clear). Let P be provisionally categorical. Then, by the definition of provisional coincidence, P cannot distinguish provisionally coincident things. Suppose toward a contradiction that P is not genuinely categorical. Then there are α and W such that α possesses P in W, but its possession of P in W depends on what worlds (other than W) it inhabits or on its genuinely categorical properties in those worlds. Thus, if there were something genuinely coincident with α in W, but differing from α in the worlds (other than W) it inhabited, or the genuinely categorical properties it had in them, that thing would lack P in W. Specifically, something existing in worlds W, W', W'' . . . , and possessing the genuinely categorical properties Y in W (Y is the set of α's genuinely categorical properties in W), Y' in W', Y'' in W'' . . . , would lack P in W. To produce such a thing, let γ ($=\alpha$), γ', γ'' . . . be entities satisfying Y, Y', Y'', . . . in W, W', W''. . . . If Z, Z', Z'' . . . are the sets of provisionally categorical properties possessed by γ, γ', γ'', . . . in W, W', W'' . . . , then, by fullness (see the preceding note), there is a β existing in exactly W, W', W'' . . . , and possessing Z in W, Z' in W', Z'' in W''. . . . Since a thing's provisionally categorical properties include its genuinely categorical properties, β meets the conditions laid down above for lacking P in W. But, by its definition, β is provisionally coincident with α in W. Since P distinguishes provisional coincidents, it is not provisionally categorical after all. QED.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

isolated. And with the categorical properties in hand, contingent identity is at last explicable: things are contingently identical in a world if they have the same categorical properties there.

Trivial cases aside, things contingently identical will not be identical as a matter of necessity. But then what about the argument that purported to show that identities obtained necessarily if at all? Was the argument invalid? No; it showed that *something* was impossible. The question is, was the refuted possibility really that of contingent identity? Looking back at the argument, the crucial assumption was this: to be contingently identical, things have to have *all* their properties, up to and including properties of the form *necessary identity with such-and-such,* in common. If that is right, then contingent identity is, as argued, impossible. So the question is, *do* contingently identical things have to have all their properties, not only categorical but hypothetical as well, in common?

They do not. To agree that they did would be to concede the very point of contingent identity and to frustrate the clear intent of its advocates, which was that it was to be a relation *compatible* with counterfactual divergence. Understanding contingent identity as sameness of nonhypothetical properties, on the other hand, preserves its point and sustains it against the "proof" of its impossibility. Still, why did it even *seem* that contingent identity entailed absolute indiscernibility? Probably because it was taken for granted that contingently identical things were (at least) *properly* identical, only—and this was their distinction—*not necessarily so.*[20] (And the expression 'contingent identity' can certainly be faulted for encouraging this interpretation.) Admittedly, contingent identity in *this* sense is not possible.[21] But there is a better and more generous way of under-

[20] True, Smart does go out of his way to emphasize that "the brain-process doctrine asserts identity in the *strict* sense" ("Sensations and Brain Processes," *op. cit.*, p. 145). But by this he seems to mean that he is not talking about the relation that one thing bears to another when they are "time slice[s] of the same four-dimensional object" or when they are "spatially or temporally continuous" (145). Certainly there is nothing to suggest that he had in mind a contrast between "strict identity" and coincidence. What is clear is that he took "strict identity" to be a relation fully compatible with hypothetical dissimilarity.

[21] Although some philosophers would say that identity *itself* can obtain contingently. *To some extent,* such philosophers can be seen as questioning the notion that is here called "identity proper" and as taking something *roughly* analogous to what is here called "coincidence" to be all the identity there is. Since this relation, which, relative to their schemes, probably deserves to be called "identity proper," can obtain contingently, the kind(s) of contingent identity they advocate is (are) in a certain sense more radical than the kind assayed here. See, for example, Allan Gibbard, "Contingent Identity," *Journal of Philosophical Logic,* IV, 2 (May 1975): 187–221; David Lewis, "Counterpart Theory and Quantified Modal Logic," this JOURNAL, LXV, 5 (March 7, 1968): 113–126, and *Counterfactuals* (Cambridge, Mass.: Harvard, 1973); and Robert Stalnaker, "Counterparts and Identity," *Mid-*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

standing contingent identity: strict and contingent identity are different relations, and, because of their differences as relations, one can obtain contingently whereas the other cannot. It only remains to spell out the formal details.

V. MODELING CONTINGENT IDENTITY

Formally, a property P is *categorical* iff necessarily, if α and β are related by refinement, then α has P iff β does, i.e., if

$$(\forall W)(\forall \alpha, \beta \epsilon \mathcal{D}(W))(\alpha \leq \beta \Rightarrow (\alpha \epsilon P(W) \Leftrightarrow \beta \epsilon P(W))).$$

What is the relation between the cumulative properties and the categorical properties? Closure implies a partial answer.

Prop. 2 If Ω is closed, then every cumulative property is categorical.
Proof: Let $P \epsilon X$, and let $\alpha, \beta \epsilon \mathcal{D}(W)$, $\alpha \leq \beta$. By u-closure, there are α^* and β^* in $\mathcal{D}(W)$ such that $\forall P \epsilon X(\alpha \epsilon P[W] \Rightarrow \alpha^* \epsilon P^{\Box}[W])$ and $\forall P \epsilon X(\beta \epsilon P[W] \Rightarrow \beta^* \epsilon P^{\Box}[W])$. We show that $\alpha \epsilon P[W] \Leftrightarrow \beta \epsilon P[W]$. Since $\alpha \leq \beta \leq \beta^*$, $\alpha \leq \beta^*$. Therefore, $\beta \epsilon P[W] \Rightarrow \beta^* \epsilon P^{\Box}[W] \Leftrightarrow \alpha \epsilon P[W]$ (by d-closure). For the converse, notice first that $\beta \leq \alpha^*$. For if $Q \epsilon \mathbb{E}_X(\beta)$, then, by d-closure, $\alpha \epsilon Q[W]$; whence $Q \epsilon \mathbb{E}_X(\alpha^*)$. Since $\beta \leq \alpha^*$, $\alpha \epsilon P[W] \Rightarrow \alpha^* \epsilon P^{\Box}[W] \Rightarrow \beta \epsilon P[W]$ (by d-closure). □

Things are *coincident* in W—written $\alpha \approx_W \beta$—if they have the same categorical properties there. But, for the definition of categoricity to achieve its purpose, the system of coincidents has to be *full* or *complete*. Informally, this means that every point in the logical space of possible coincidents must actually be occupied; formally, for any (partial) function f from worlds W to things existing in W, there is a thing existing, and coincident with $f(W)$, in exactly the worlds in f's domain. Ω is *full* if it satisfies

(F) $\forall f{:}W \epsilon \mathcal{W} \rightarrow f(W) \epsilon \mathcal{D}(W)$
$\exists \alpha\ [\mathcal{W}(\alpha) = \mathrm{dom}(f)\ \&\ \forall W \epsilon \mathcal{W}(\alpha)\ \alpha \approx_W f(W)]$

Prop. 2 showed that, if Ω is closed, then every cumulative property is categorical. For the converse, let Ω be *maximal closed* if

(M) $\Omega = \langle \mathcal{W}, X \rangle$ is closed, and there is no X' extending X such that $\langle \mathcal{W}, X' \rangle$ is closed.

If Ω is maximal closed and full, then every categorical property is cumulative. In other words, a property is cumulative if and only if it is categorical.

---

west Studies in Philosophy, XI (Minneapolis: Minnesota UP, 1986). (Take note: as a descriptive account of these authors' theories, which are anyway not very similar to one another, the above is not reliable.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Prop. 3 Let Ω be maximal closed and full. Then

$$\forall P\ [P \text{ is cumulative} \Leftrightarrow P \text{ is categorical}]$$

Proof: [$\Rightarrow$] This is just Prop. 2. [$\Leftarrow$] Let P be categorical. I claim that $\langle \mathcal{W}, X + P \rangle$ is closed. For u-closure, let $Z \subseteq X$ and suppose that α has $Z + P$ in W. By fullness, there is an α^* which exists in W only and which is coincident with respect to (X) with α in W. Since α^* exists in W only, it has all its properties essentially. In particular, α^* has all of α's categorical (w.r.t. X) properties essentially, and so (by Prop. 2) α^* refines α (w.r.t. X). Since P is categorical (w.r.t. X), α^* has P in W too, and so it has P essentially in W. Thus α^* has $(Z + P)^{\Box}$ in W. For d-closure, let $Z \subseteq X$ and let $\alpha \leq \beta \epsilon (Z + P)^{\Box}[W]$. By the d-closure of Ω, $\alpha \epsilon Z[W]$, and, since P is categorical (w.r.t. X), $\alpha \epsilon P[W]$ too. Since $\langle \mathcal{W}, X + P \rangle$ is both u-closed and d-closed, P is in X.[22] $\Box$

VI. ESSENCE AND IDENTITY

Can distinct things have the same cumulative essence? So far, nothing prevents it. For example, there is nothing to rule out the following: there are exactly two possible worlds, W and W'; α and β exist in both worlds; γ exists in W alone; and γ' exists in W' alone. In this situation γ will have to refine both α and β in W, and γ' will refine both α and β in W'. From the fact that α and β have a common refinement in each world, it quickly follows that they are coincident, i.e., have the same categorical properties, in each world. But, by Prop. 3, the categorical properties are exactly the cumulative properties; and, if α and β share their cumulative properties in every world, how can they have different cumulative essences?

Actually, this raises the critical question, avoided until now, of how coincidence and identity are related. Can α and β exist in the same worlds, be coincident in all of them, and still be distinct? It is hard to imagine how they could. For, presumably, distinct items differ in one or another of two ways. Either they exist in different worlds, or they exist in the same worlds and are unlike, i.e., have different categori-

[22] Although it is wrong to try to *explicate* contingent identity in terms of identity of modal states (the notion of modal state depending on that of contingent identity), Prop. 3 suggests a way to define modal states so that the equation comes out true. Call α^* α's *state* in W iff α^*'s cumulative essence is exactly the set of α's categorical properties in W. To see that if α exists in W its state in W exists too: By fullness, there is a β existing in W alone and coincident with α there. Thus β's categorical properties in W are exactly α's. By Prop. 3, β's cumulative properties in W are exactly α's categorical properties in W. Since β has all its properties essentially, β's cumulative essence is the set of α's categorical properties in W. Now it is easy to verify that things are coincident in a world iff they have strictly identical states there: α coincides with β in W iff α and β have the same categorical properties in W iff anything whose cumulative essence is the set of α's categorical properties is also something whose cumulative essence is the set of β's categorical properties iff any state of α in W is a state of β in W. (For the uniqueness of α's state in W, see Prop. 4.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

cal properties, in at least one of them. Between distinct things, that is, there have got to be either intra-world or extra-world differences.

But what is the argument for this? If things exist in the same worlds, then, unless they have different categorical properties in at least one of them, the hypothesis of their distinctness can find no foothold. Take the standard example of "indiscernible" spheres afloat in otherwise empty space (suppose for argument's sake that they exist in no other world). If the spheres were in *exactly the same place,* could they still be reckoned distinct? An hypothesis *so* gratuitous is beyond not only our powers of belief, but even our powers of stipulation. If, on the other hand, the spheres are in different places, then they differ on the (presumably categorical) properties of being in those places. (The properties have to be different, because they map the world in question to different spheres.)[23]

Call a property-model *separable* if it satisfies

$$(5) \quad \forall\alpha\, \forall\beta\, [(\mathcal{W}(\alpha) = \mathcal{W}(\beta)\ \&\ \forall W \epsilon \mathcal{W}(\alpha) \cup \mathcal{W}(\beta)\ \alpha \approx_W \beta) \Rightarrow \alpha = \beta]$$

The last proposition shows that, if Ω is (besides being closed) separable, then things with the same cumulative essence are identical.

Prop. 4 Let Ω be closed and separable. Then

$$\forall\alpha\, \forall\beta\, [\mathbb{E}_X(\alpha) = \mathbb{E}_X(\beta) \Rightarrow \alpha = \beta]$$

Proof: Let α and β have the same X-essence. Since α and β X-refine each other, by Prop. 1 they exist in the same worlds. By the definition of X-categoricity, in each of these worlds any X-categorical property attaching to either attaches also to the other. Thus α and β have the same X-categorical properties, and so coincide, in every world where they exist. By separability, α is identical with β. □

To this extent, essence determines identity.[24]

[23] Granted, one cannot identify these properties without appeal to the objects that have them; but the claim was that they have different categorical properties, not that one can distinguish them by their categorical properties. Granted also, except in connection with the world in question, the properties' extensions will be to a large extent arbitrary (when is something in this world in the same place as the first sphere in that one?); but that does not matter, so long as the arbitrary choices are made in such a way that the resulting properties are categorical.

[24] Even if this result is accepted, there is plenty of room for doubt about its precise significance. Briefly, one worry is that to distinguish α's cumulative essence from β's, one would already have to be able to distinguish α from β. But this seems to confuse the metaphysical thesis that distinct things have different essences with the epistemological doctrine that distinct things can always be *distinguished* by their essences. Second, not everything here called a "property"—basically, functions from worlds to extensions—is a *genuine property.* If this means that genuine properties are not functions, this is granted; but it does not matter, if, for every

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

VII. APPLICATIONS

(A) Treating contingent identity as sameness of categorical properties goes *part* of the way toward solving a problem David Wiggins raises for relative identity in *Sameness and Substance*. He argues there that, since (i) what sets identity relations apart from the common run of equivalence relations is their satisfaction of Leibniz's law, and (ii) no variety of relative identity can satisfy (an unrestricted version of) Leibniz's law, (iii) relations of relative identity are not identity relations (pending discovery of a suitably restricted form of Leibniz's law). To answer this argument, one would need an uncontrived law of the form: if α is the same f as β, then α and β have thus-and-such properties in common. However, Wiggins thinks such a law will prove impossible to formulate:

> It seems that the very least we shall require is more information about the case of the *restricted* congruence that results from the g-identity, for *some one* sortal concept g, of x and y. No stable formulation of restricted congruence is available, however. Nor, I suspect, will it ever be given (39).

But, if contingent identicals are seen as the *same concrete thing*, then a rigorous restriction of Leibniz's law is at hand: if α and β are the same concrete thing, then they have the same categorical properties.[25]

(B) Nearly everyone's gut reaction to functionalism is that *phenomenal* properties, at any rate, cannot be functional, because nothing functional can attain to the "manifest" character of felt experience. Perhaps this idea finds support in the categorical/hypothetical distinction. The property of playing functional role R is the property of bearing certain complicated counterfactual relations to inputs, outputs, and the players of various other functional roles. Details aside, such a property is obviously hypothetical. But the property of painfulness (note: not the property of causing pain, but that of being pain) *seems* to be a categorical property *par excellence*.

---

world-to-extension function, there is some genuine property such that the function takes each world to the set of things possessing that property therein. But the criticism survives in the form: not all world-to-extension functions (not even all cumulative ones) *are* induced in this way by genuine properties. And that is undeniable. Further progress depends on figuring out what makes genuine properties genuine. (For more on the difference between genuine and pseudo properties, see the articles by Sydney Shoemaker mentioned in note 15 above.)

[25] Whether this helps with the general problem of saying what properties relative identicals must have in common is another question, but one perhaps worth exploring. See also Nicholas Griffin, *Relative Identity* (New York: Oxford, 1975), secs. 1.2 and 8.5.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Therefore, painfulness is not a functional property. (Note that this affects only the version of functionalism that flatly identifies mental properties with functional properties.)

(C) That mental and physical events are not properly identical is argued not only by their essential differences (emphasized by Kripke), but also by their causal differences. Suppose Smith's pain is identical with neural event ν, which causes neural event ε. Then the strict identity theorist will have to say that Smith's pain caused ε; but that seems questionable, because ν's *being* Smith's pain contributed nothing to its production of ε (one wants to say: even if ν had not been Smith's, or any, pain, ε would still have eventuated). If Smith's pain and ν are only coincident, on the other hand, then *naturally* they will have different causal powers and susceptibilities; and a sensitive counterfactual theory of causation might be able to *exploit* their essential differences to predict their causal differences.[26] Irrelevant qualifications aside, an event α causes an event β only if it is *required* for β, in the following sense: given any (actually occurring) event $\gamma \approx \alpha$ whose essence does not include α's, if γ had occurred in α's absence, β would not have occurred; and only if it is *enough* for β, in the following sense: given any (actually occurring) event $\gamma \approx \alpha$ whose essence is not included in α's, γ is not required for β. As for Smith's pain (call it π), that ε would have occurred even in π's absence, provided that ν had still occurred, shows that π is not required for ε; and that ν is (let us assume) required for ε shows that π is not enough for ε either. Complementary considerations show how it can be Smith's decision, rather than the corresponding neural event, that causes her action.

STEPHEN YABLO

University of Michigan/Ann Arbor

[26] Thanks to Barry Loewer for talking to me about this; he and Paul Boghossian are working along similar lines.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

VAN FRAASSEN ON EXPLANATION*

THERE should be no doubt about the fact that Bas van Fraassen has made substantial contributions to our current understanding of scientific explanation. But we believe that there is reason for doubt as to exactly what the contributions are. Chapter V of *The Scientific Image*, "The Pragmatics of Explanation," offers the most detailed account of van Fraassen's view of explanation. We find both the title and the view ambiguous. The purpose of the present discussion is to underscore the difference between a theory of the pragmatics of explanation and a pragmatic theory of explanation. We believe that van Fraassen has offered the best theory of the pragmatics of explanation to date, but we shall argue that, if his proposal is seen as a pragmatic theory of explanation, then it faces serious difficulties.

I

Before we turn to van Fraassen's positive views, we want to consider his response to the tradition of theorizing about explanation. According to van Fraassen, there are two main problems "of the philosophical theory of explanation." These are "to account for legitimate rejections of explanation requests, and for the asymmetries of explanation" (146). Van Fraassen's solution to the former problem seems to us to be ingenious and substantially correct (see below). His treatment of the asymmetries of explanation we find deeply puzzling.

Within the mainstream of philosophical reflection about explanation, the problem of asymmetries arises because there are arguments which are closely related, which accord equally well with the conditions set down by models of explanation, and which differ dramatically in their explanatory worth. For present purposes assume either that some explanations (including the examples to be considered) are arguments or that some arguments (including those to be considered) provide a basis for acts of explanation.[1] Then the challenge is

* The present paper grew out of discussions between us at the Minnesota Center for the Philosophy of Science during the fall of 1985. We would like to acknowledge the support of the National Endowment for the Humanities for a grant that made our collaboration possible. We are also grateful to Bas van Fraassen for some helpful clarifications of his ideas, made in response to an earlier draft. Parenthetical page references are to *The Scientific Image* (New York: Oxford, 1980).

[1] The idea that arguments might provide the basis for acts of explanation is suggested in Philip Kitcher, "Explanatory Unification," *Philosophy of Science*, XLVIII, 4 (December 1981): 507–531; and is articulated in some detail by Peter Railton, "Probability, Explanation, and Information," *Synthese*, XLVIII, 2 (August 1981): 233–256. Although one of us (WCS) rejects the thesis that explanations are

0022-362X/87/8406/0315$01.60 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to differentiate between the argument that derives the length of a shadow from the height of a tower, the elevation of the sun, and the principles of optics and the argument that derives the height of the tower from the length of the shadow, the elevation of the sun, and the principles of optics. The former seems to be (a potential basis for) an explanation, whereas the latter does not.

One line of solution, hinted at by Carl Hempel in discussion of an analogous case,[2] is to propose that there is no real difference between the two arguments and that the feeling of difference arises from anthropomorphic ideas from which we ought to liberate ourselves. This is not very convincing, and van Fraassen appears to adopt a more satisfactory method of dissolving the problem. *One* way to understand his fable, "The Tower and the Shadow," is as an attempt to show that the claim of explanatory difference is shortsighted. Failing to appreciate that arguments are explanations (the basis of explanations) only relative to context, we assess the explanatory merits of the derivations by tacitly supposing contexts that occur in everyday life. With a little imagination, we can see that there are alternative contexts in which the argument we dismiss would count as explanatory.

In van Fraassen's story, a character offers the following explanation of the height of a tower:

> That tower marks the spot where [the Chevalier] killed the maid with whom he had been in love to the point of madness. And the height of the tower? He vowed that shadow would cover the terrace where he first proclaimed his love, with every setting sun—that is why the tower had to be so high (133/4).

Now we grant that van Fraassen's story describes a context in which the utterance of these words constitutes an explanation for the position and height of the tower. But this will solve the *traditional* problem of the asymmetries of explanation only if one can claim that the argument underlying the quoted passage is the argument that the unimaginative have dismissed as nonexplanatory.

It seems obvious that this is not so. For the (D-N; deductive-nomological) argument that provides the basis for the act of explanation van Fraassen relates does not take the form of deducing the

---

arguments, or involve arguments in any essential way, this issue does not affect the present discussion of van Fraassen's views in any significant fashion.

[2] "Aspects of Scientific Explanation," in *Aspects of Scientific Explanation* (New York: The Free Press, 1965). See pp. 352/3.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

height of the tower from the length of the shadow (with the elevation of the sun and the principles of optics as the only other premises). Rather, we begin with some initial conditions about the psychological characteristics of the Chevalier—he wanted to build a tower with certain properties; he knew certain physical facts. Using general principles of rationality, we infer a statement to the effect that the Chevalier came to believe that, if he built a tower of the appropriate height on the appropriate spot, it would meet his desiderata. Using yet another principle of rationality, we infer that the Chevalier built the tower to these specifications, and, using background principles about the stability of the height and position of such large physical objects, we conclude that the tower has the height and position it has.

It appears that an obvious way to interpret van Fraassen is mistaken: his story does not provide a context in which an argument wrongly dismissed as explanatory shows its explanatory worth. Moreover, since van Fraassen points out, quite explicitly, the dependency on desires (132), we take him to appreciate that his story does not solve the traditional problem of the asymmetries of explanation. Instead, we construe him as claiming that the problem as we have posed it—a problem that talks about arguments and their merits as explanations (the bases of explanations)—is *mis*posed. Once the topic is approached in terms of van Fraassen's favored pragmatic machinery, we are to see that an *answer* that we might have considered inappropriate can have explanatory worth in the right context.

But this leaves us with puzzles. If we cannot formulate the traditional problem of the asymmetries of explanation in terms of arguments, then how is the problem to be formulated? Does an analogous problem arise within van Fraassen's own theory? Is it resolved by that theory? We shall return to these questions below.

II

According to van Fraassen, an explanation is an answer to a question Q of the form: Why P_k? where P_k states the fact to be explained—i.e., the explanandum (phenomenon). Any such question can be identified as an ordered triple $\langle P_k, X, R\rangle$, where P_k is called "the topic" of the question, $X = \{P_1, \ldots, P_k, \ldots\}$ is its contrast class, and R is its relevance relation. Such a question is posed in a context that includes a body of background knowledge K. Q also has a *presupposition*, namely

(a) P_k is true;
(b) each P_j in X is false if $j \neq k$;
(c) there is at least one true proposition A that bears relation R to $\langle P_k, X\rangle$.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(a) and (b) together constitute the *central presupposition* of Q. The why-question Q *arises* in the given context if K entails the central presupposition of Q and does not entail the falsity of (c). That is, it is altogether appropriate to raise Q even if we do not know whether there is a direct answer or not, provided the central presupposition is fulfilled.

If the question does not arise in the context, it should be rejected rather than answered directly. This can be done by offering a corrective answer, i.e., a denial of one or more parts of the presupposition. If the central presupposition is satisfied but (c) is in doubt, a corrective answer to the effect that (c) is false may be suitable.

If the question arises in the given context, it is normally appropriate to provide a *direct answer*. The canonical form of a direct answer to Q is

(*) P_k in contrast to the rest of X because A.

The following conditions must be met:

(i) A is true.
(ii) P_k is true.
(iii) No member of X other than P_k is true.
(iv) A bears R to $\langle P_k, X\rangle$.

A is the *core* of the answer, for the answer can be abbreviated "Because A."

Since, typically, the person S_q who asks the question Q might be someone with a different body of knowledge from the respondent S_r, we might be tempted to say that two different contexts are involved. It seems more in keeping with van Fraassen's approach, however, to understand that S_q and S_r are operating in a common context with a common body of background knowledge K determined roughly by the state of science at the time. Thus, K may contain many propositions that neither the questioner nor the respondent knows. Moreover, S_q may have false beliefs that are in conflict with propositions in K. S_r may therefore offer corrective answers to flawed questions by pointing to items in K.

Whether A, the core of the answer to Q, is relevant depends solely on the relevance relation R. If A bears R to $\langle P_k, X\rangle$ then A is, by definition, the core of a relevant answer to Q. This way of stating the matter raises a difficulty. In his informal remarks, van Fraassen repeatedly refers to R as a "relevance relation," but he incorporates no relevance requirement on R in the formal characterization. Indeed, he points to the absence of any problematic constraint that would

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

seek to capture "the inextricably modal or counterfactual element" (143). Now, if R happens to be a relevance relation, then it is indeed correct to say that A is relevant to $\langle P_k, X\rangle$. But, as we shall now show, the lack of any constraints on "relevance" relations allows just about anything to count as the answer to just about any question.

III

Let P_k be any true proposition. Let X be any set of propositions such that P_k belongs to X and every member of X apart from P_k is false. Let A be any true proposition. Let R be $\{\langle A, \langle P_k, X\rangle\rangle\} \cup S$, where S is any set of ordered pairs $\langle Y, Z\rangle$ such that Y is a proposition and Z is $\langle V, W\rangle$ where V is a proposition and W a set of propositions, one of whose members is V. Then there is a why-question $Q = \langle P_k, X, R\rangle$, and A is the core of a direct answer to Q. Moreover, it is easy to see that, with suitable restrictions on S (i.e., that S contain no $\langle Y, Z\rangle$ such that Y is true and Z is $\langle P_k, X\rangle$), A may be the core of the only direct answer to Q. Hence, for any true propositions P_k and A, there is a why-question with topic P_k such that A is the core of the only direct answer to that question. If explanations are answers to why-questions, then it follows that, for any pair of true propositions, there is a context in which the first is the (core of the) only explanation of the second.

We take it that this result is counterintuitive. Indeed, we would view it as a *reductio* of van Fraassen's account of explanation. How can it be avoided?

One way of blocking the trivialization we have outlined would be to impose restrictions on relevance relations. We shall consider this possibility below. First, let us note that van Fraassen's theory of explanation comes in two parts: there is a thesis about what answers to why-questions are, and there is a thesis about how to evaluate answers to why-questions. Perhaps we can use the latter part of the story to defend against the trivialization that threatens the former.

According to van Fraassen, we evaluate answers to why-questions on three different grounds. We ask whether those answers are probable in light of our knowledge, we ask whether they *favor* the topic against the other members of the contrast class, and we ask whether they are made wholly or partially irrelevant by other answers that could be given. Using a notion that van Fraassen often employs in his informal remarks but does not define, let us say that an answer is *telling* if it scores well according to these criteria. More exactly, let us propose that an answer is more or less telling according to its performance on the three criteria. We shall be most interested in maximally telling answers. We shall call them *perfect* answers.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Notice that the theory of evaluating answers to why-questions allows us to compare different answers to the same questions. It does not enable us to assess the degree to which an answer to one question is more telling than an answer to another question. If the questions are of the contrived kind that we introduced at the beginning of this section, then there will be no more telling answer to them than the contrived answer. However, we may easily introduce a grading of questions by considering whether they admit of answers that favor their topic.

Let us say that questions are more or less *well-founded* to the extent that they admit of telling answers. Thus, a question will be *maximally well-founded* if it admits of a perfect answer. Suppose now that P_k is any true proposition, A any proposition, and X any set of two or more propositions such that P_k is its only true member. Let K be a set of propositions that includes both P_k and A, as well as the negations of all the other propositions in X. Then, we claim, there is a why-question whose topic is P_k, whose contrast class is X, such that A is an essential part of a perfect answer to that why-question.

To demonstrate this we need to examine in somewhat more detail van Fraassen's criteria for evaluating answers. On the first criterion, we award high marks to answers if they receive high probability in light of our background knowledge. A corollary of this is that, if the answer belongs to our background knowledge—as is often the case when we give scientific explanations—then it does as well as possible according to this criterion.

The second criterion (favoring) is less straightforward. Van Fraassen's idea is that the answer, to score well, should increase the distance between the probability of the topic and the probabilities of the other members of the contrast class. Typically, the answer *alone* will not redistribute probabilities in this way. Rather, the answer, taken together with certain auxiliary information, will redistribute the probabilities. However, we cannot suggest that the answer plus the total background knowledge K achieves this result; for, in cases where the topic and the negations of the other members of the contrast class belong to K, the suggestion would lead to immediate trivialization. Van Fraassen therefore suggests that the redistribution of probabilities be achieved by the answer in conjunction with "a certain part $K(Q)$ of K," where $K(Q)$ is supposed to be contextually determined.

We need not delve into the problems of deciding exactly what counts as singling out the topic within the contrast class, since we shall use a case in which van Fraassen explicitly recognizes that an

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

answer is maximally successful. He writes: "If $K(Q)$ plus A implies B and implies the falsity of $C, \ldots, N$ then A receives in this context the highest marks for favoring the topic B" (147). [There is a switch in notation here; the topic is B, the contrast class is $\{B, C, \ldots, N\}$].

Van Fraassen's third criterion concerns the availability of superior answers. The answer A loses marks if it has a rival that fares better, perhaps because the rival receives higher probability in light of background knowledge K, perhaps because the rival favors the topic more than A does, perhaps because the rival screens off A from the topic. Now A does not have to fear any rival if A belongs to K and if A plus $K(Q)$ implies the topic and the negations of the other members of the contrast class. For, under these circumstances, no rival can be more probable in light of K, no rival can do better at favoring the topic, and no rival can screen A off from the topic.

We conclude that any A belonging to K which, in conjunction with $K(Q)$, implies the topic P_k is a perfect answer to the question $\langle P_k, X, R\rangle$ (provided, of course, that it is an answer to this question).

Let us therefore define a "relevance relation" R as follows: we stipulate that R holds between B and $\langle P_k, X\rangle$ just in case P_k is a logical consequence of B. Let Z be the disjunction of all the propositions in X apart from P_k, and let B be the proposition

$$A \cdot (A \supset P_k) \cdot \sim Z$$

This proposition bears R to $\langle P_k, X\rangle$, and hence it counts as the core of a direct answer to the why-question $\langle P_k, X, R\rangle$. Moreover, by our earlier assumption, A, P_k, and $\sim Z$ belong to K. This means that all the conjuncts in B, and hence B itself, belong to K. Thus, B will be completely successful according to van Fraassen's first criterion for evaluating answers. Because $P_k \cdot \sim Z$ is a logical consequence of B, B maximally favors P_k—and we do not need to worry about how $K(Q)$ is selected since $P_k \cdot \sim Z$ is a consequence of B alone. Finally, because of this implication, there is no reason to fear that B will be screened off by some rival answer. Therefore, B is a perfect answer to $\langle P_k, X, R\rangle$.

We have devised one way of finding, for any pair of true propositions A, P_k, a why-question with P_k as topic to which there is a perfect answer with A as an essential part of its core. Moreover, once we see how the construction we have given is possible, it is easy to generate variations on the same theme. For example, if van Fraassen's account does not contain context-independent principles that preclude the possibility of assigning $(A \supset P_k) \cdot \sim Z$ to $K(Q)$, then it will be possible

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to claim that A is the core of a direct perfect answer to some question with P_k as topic.

We conclude that the machinery that van Fraassen introduces in his discussion of the evaluation of answers does not avail in protecting him against the kind of trivialization we presented at the beginning of this section. The moral is that, unless he imposes some conditions on relevance relations, his theory is committed to the result that almost anything can explain almost anything. Some kinds of relations R are silly, and why-questions that embody them are silly questions. If we pose silly questions, we should not be surprised to get silly answers.

IV

Let us now consider a concrete example. Suppose S_q asks why John F. Kennedy died on November 22, 1963, where

P_k = JFK died 11/22/63, X = {JFK died 1/1/63, JFK died 1/2/63, . . . , JFK died 12/31/63, JFK survived 1963},

and R is a relation of astral influence. (One way to define R is to consider ordered pairs of descriptions of the positions of stars and planets at the time of a person's birth and propositions about that person's fate.) An answer with core A might consist of a *true* description of the positions of the stars and planets at the time of JFK's birth. Moreover, using astrological theory as background, one might be able to infer (at least with high probability) that JFK would die on 11/22/63.

We suggest that, in the context of twentieth-century science, the appropriate response to the question is rejection. According to our present lights, astral influence is not a relevance relation. We believe that the positions of the stars and planets on JFK's birthday have no effect on the probability of death on any particular day. Adding the knowledge of those positions does nothing to redistribute the probabilities of death among the members of the contrast class. The moral we draw here—as in the last section—is that van Fraassen's conditions (a)–(c) on answers to why-questions need to be supplemented by adding

(d) R is a relevance relation.

Moreover, we claim that (d) cannot be analyzed simply in terms of demanding that, if A bears R to $\langle P_k, X\rangle$, then A must redistribute the probabilities on X. For we can meet that demand by considering the proposition

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

B = A · (If A, then JFK died on 11/22/63) · JFK did not die on 1/1/63 · . . . JFK did not die on 11/21/63 · JFK did not die on 11/23/63; · . . . · JFK did not survive 1963)

and defining the relation of astral influence R so that R contains $\langle B, \langle P_k, X\rangle\rangle$.

Once again, let us consider the question from the perspective of van Fraassen's account of evaluating answers. We note, first, that the true description of the positions of stars and planets at JFK's birth accords with our current scientific knowledge. So the answer gets high marks on this score. Second, we ask to what extent A favors P_k vis-à-vis the other members of X. On this criterion A fares poorly (although B, of course, does not). Perhaps an answer that negatively favored the topic might get still lower marks—though it is not clear to us that it should, since discovering a relevant factor seems better than offering an irrelevancy. Third, we must compare A with other answers to Q. This criterion has three parts. (i) Since A is true and since it belongs to our body of knowledge, no other answer can be more probable. (ii) Since no astrological answer is relevant, all astrological answers equally fail to favor the topic. (iii) Since every astrological answer is irrelevant, screening off is beside the point.

The result is that A is not telling. There is no telling answer to our original question. If we amend the question, we can produce a relative to which B is a maximally telling answer. In our view, both the questions ought to be rejected, and van Fraassen needs to supplement his theory of explanation with an account of relevance relations.

The astrological answer has a further twist, however. As van Fraassen explains, our general background knowledge K—suitably restricted to $K(Q)$ to avoid trivialization—furnishes a *prior distribution* of probabilities over the contrast class X. (Note that, in discussing favoring with respect to the contrived example of section III, and answer B of this section, we were entitled to take $K(Q)$ to be any subset of K because we had no need of any additional premises in generating the most extreme distribution of probabilities over the contrast class.) Given A, we have a *posterior distribution* of probabilities over X. (It should *not* be assumed that the prior distribution assigns equal probabilities to all members of X; surely, survival beyond 1963 was antecedently more likely than death on any given day, and surely some days are more dangerous than others in the life of a U.S. President.) A is the core of a relevant answer to Q only if addition of A to $K(Q)$ would yield a posterior distribution different from the prior distribution. But what sorts of probabilities are these?

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

If they are S_q's personal probabilities, then, given that S_q is a believer in astrology, we might well expect that knowledge of A would lead to a different distribution. So A would be relevant after all.

Van Fraassen might reply that astrological answers are debarred by his (frequently repeated)[3] remarks that explanations make use of accepted scientific theories. The astrological answers are precluded by the fact that they contain statements that are inconsistent with the background knowledge K. But this seems to mistake the purport of our examples. The statement A *belongs* to the background corpus; it is simply a report of the positions of the heavenly bodies at the time of JFK's birth, and we can assume that this report derives from the best current science. Of course, if the answer includes further bits of astrological theory designed to connect A with the statement that JFK died on 11/22/63, then van Fraassen will have grounds for ruling it out. But, if the favoring of the topic is achieved solely through S_q's personal probabilities, then there is nothing in the answer to which van Fraassen can point as defective. Similarly, there is nothing in B that would be debarred on the basis of an appeal to background knowledge, for all the statements in B belong to the background corpus.

It should now be clear that these examples work by exploiting the laxity of the conditions on the relevance relation in order to reintroduce "explanations" that van Fraassen hopes to debar by emphasizing the idea that good explanation must use good science. Unless there are constraints on genuine relevance relations, we can mimic the appeal to deviant beliefs in giving pseudoexplanations by employing deviant relevance relations. Hence, if van Fraassen is serious in his idea that genuine explanations must not make appeal to "old wives' tales," then he ought to be equally serious about showing that relevance is not completely determined by subjective factors. If we are talking about distributions and redistributions of personal probabilities, they must be subject to some kinds of standards or criteria.

[3] At the outset of his exposition of his theory of why-questions, van Fraassen remarks, "This evaluation [of answers] proceeds with reference to the part of science accepted as 'background theory' in that context" (141). Earlier, he had remarked that, "To ask that . . . explanations be scientific is only to ask that they rely on scientific theories and experimentation, not on old wives' tales" (129), and "To sum up: no factor is explanatorily relevant unless it is scientifically relevant; and among the scientifically relevant factors, context determines explanatorily relevant ones" (126). In conclusion, he says, "To call an explanation scientific is to say nothing about its form or the sort of information adduced, but only that the explanation draws on science to get this information (at least to some extent) and, more importantly, that the criteria of evaluation of how good an explanation it is, are being employed using a scientific theory" (155/6).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Coherence is one such criterion, but it cannot be sufficient. To be scientifically acceptable, the redistribution of probabilities must involve differences in objective probabilities (frequencies or propensities) in some fashion.

V

When van Fraassen explicitly discusses kinds of relevance relations, the kinds he picks out are fairly familiar from the literature on scientific explanation: we discover such relations as physical necessitation, being etiologically relevant, fulfilling a function, statistical relevance, and, in the fable of "The Tower and the Shadow," a relation of intentional relevance. We have been arguing that there are some relations that ought not to be allowed in *any* context as genuine relevance relations. Thus, there appears to be a distinction to be drawn between the relations that can serve, in some context or another, as relevance relations (paradigmatically those relations which figure in van Fraassen's discussions) and those which cannot (such as the contrived relations of the last two sections).

How the distinction should be drawn depends on a very general issue about scientific explanation. Is there a set of genuine relevance relations that underlie the genuine why-questions for all sciences and for all times? Those who give an affirmative answer will see a full theory of explanation as offering a specification of the kinds of relevance relations that may underlie genuine why-questions. That specification would be strongly context-independent in that it would pick out the candidates for any given context of posing a why-question, and the candidates would always be the same.

But perhaps there is no such invariant set of genuine relevance relations. The set of genuine relevance relations may itself be a function of the branch of science and of the stage of its development. Consider the abandonment of teleological explanations in physics after the scientific revolution. This can be viewed as a modification of the set of relevance relations: in the context of Aristotelian physics, the notion of teleological relevance was a genuine relevance relation, in the context of Newtonian physics it lost this status. Uniformitarians (those sympathetic to the view of the last paragraph) will deplore this relativism, contending that the notion of teleological relevance never was a genuine relevance relation and that its status was exposed during the scientific revolution. They will accuse relativists of confusing the variation in beliefs about relevance with the relativity of relevance itself.

We do not need to settle this dispute because, on both accounts, there is a nontrivial task of distinguishing genuine relevance relations

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

from the contrived relations of the last two sections. Just as pluralists about literary works will insist that there are many interpretations of *Hamlet* while denying that any reader's fancy counts as an interpretation, so too, relativists should concede that there are some relations that are not genuine relevance relations at any historical stage of any science. The most thorough version of a relativist account of explanation would consist in specifying those principles which determine, for each historical stage of each science, the selection of certain relations as genuine relevance relations. A more modest (and more sensible) approach would be to consider some particular science (or sciences) throughout some particular period and to identify the pertinent relevance relations. Thus, one might focus on contemporary physics and try to distinguish the associated genuine relevance relations from the residue of relations—the contrived, the discarded, and so forth.

Uniformitarians, ambitious relativists, and modest relativists all face the same kind of task. Although we do not know which version of the task he would wish to undertake, van Fraassen has remarked to us (in conversation) that he recognizes the importance of distinguishing genuine relevance relations and that he takes Aristotle's list of types of causes to be a promising start on drawing the distinction. We now want to suggest that completion of the task will require that van Fraassen solve most (if not all) of the traditional problems that have beset theories of explanation. For, depending on one's commitment on the large issue we have left unresolved, these problems take the form of showing why certain relations do not belong to the single set of genuine relevance relations which is associated with all sciences at all times, or of showing why certain relations do not belong to any of the sets of genuine relevance relations associated with different sciences at different times, or of showing, for some particular science(s) and period of interest, why certain relations do not belong to the associated set of genuine relevance relations. Henceforth, we intend that our presentations of problems should be systematically ambiguous among these forms.[4]

To simplify matters, we shall confine our attention to difficulties that arise in what Hempel would have viewed as deductive explanation. Consider the simple relation of derivation. This relation holds

[4] We are very grateful to an editor of this JOURNAL who raised a question which prompted us explicitly to distinguish these three ways of pursuing the theory of explanation and, thus, substantially to improve on the formulations of an earlier draft.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

between A and $\langle P_k, X\rangle$ just in case there is a (first-order) derivation of P_k from A plus additional premises in $K(Q)$. We can define any number of relations by imposing constraints on the kinds of statements that should figure in the premises. Thus, to recall a famous Hempelian example, let P_k be the proposition that Horace is bald, and R be the relation of Greenbury-school-board-derivation that holds between A and P_k just in case A is a conjunction of propositions one of whose conjuncts is the proposition that Horace belongs to the Greenbury school board, P_k is derivable from A, and there is no conjunct in A that could be deleted while still enabling P_k to be derivable from the result. Suppose that X includes the propositions that Horace is bald and that Horace is not bald. Let A be the proposition that Horace is a member of the Greenbury School Board and that all members of the Greenbury school board are bald. $\langle P_k, X, R\rangle$ is a van Fraassen why-question to which A is a direct answer, and a perfect answer, to boot.

We claim that the question we have just artificially constructed is not a genuine why-question and that A is no explanation of Horace's baldness. Moreover, we suggest that most (if not all) of the examples of nonexplanatory arguments that Hempel hoped to exclude—both those he succeeded in debarring and those which have caused persistent problems for the theory of D-N explanation—give rise to corresponding "relevance" relations that van Fraassen ought to exclude. As an illustration, let us return to his solution to the problem of the asymmetries of explanation in the light of what we have discovered about his treatment of why-questions.

The proposition that the tower was built on the spot where the Chevalier killed his beloved and that it was built of such a height that its shadow would fall across the terrace where he first vowed his love is relevant to the topic of the question, "Why is the height of the tower h?" if we construe the relevance relation to be that of intentional relevance. That is just another way of putting the point that there is a perfectly good Hempelian argument that derives the height of the tower from premises about the Chevalier's attitudes and from psychological laws. But, if we are moved by the traditional problem of the asymmetries of explanation, what we want to know is whether there is a context in which the statement "The length of the shadow is l" answers the question "Why is the height of the tower h?" in virtue of the fact that the assertion about shadow length, together with premises about the angle of elevation of the sun and the propagation of light [which may be relegated to the background $K(Q)$] favor the topic as against other propositions ascribing different

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

heights. For that (or something very like it) is the translation into van Fraassen's idiom of the asymmetry problem that has bedeviled Hempel and his successors.

Now, unless we impose very delicate constraints on relevance relations, it is easy to contrive a maximally well-founded question $\langle P_k, X, R\rangle$ such that the proposition ascribing shadow length will be the core of a perfect answer. The trick should be apparent by now: take P_k to be the proposition that ascribes the actual height to the tower, let X be a collection of propositions ascribing different heights, let R be the relation of *censored Hempelian derivation*—a relation that holds between A and $\langle P_k, X\rangle$ just in case there is a D-N argument that derives P_k from A plus additional premises in $K(Q)$. (Quite evidently, we could impose additional constraints so as to rule out the use of the psychological principles on which van Fraassen's account turns, and thus to ensure that the only available D-N arguments are those which invert the usual order of explanatory derivation). We take $K(Q)$ to be fixed in such a way as to include the proposition ascribing the elevation of the sun and the laws of propagation of light. This is surely quite reasonable, for some such $K(Q)$ will have to be allowed if we are to countenance the proposition that the height of the tower is h as the core of an answer to the question, Why is the length of the shadow l? So van Fraassen's theory allows explanations which correspond to those D-N explanations which intuitively "run the wrong way."

We suggest that this is a mistake. Just as the contrived questions of sections III and IV should be eliminated by the imposition of constraints on relevance relations, so too the question of the last paragraph and its accompanying perfect answer ought to be banished. For otherwise, van Fraassen's account of explanation will be deficient in exactly the way that Hempel's own treatment was. Every kind of asymmetry that arises for the D-N model can be generated within van Fraassen's framework. This means that, far from solving the problem of the asymmetries of explanation, van Fraassen presupposes a solution to that problem. Thus, if we are right, van Fraassen has offered a beautiful treatment of the pragmatics of explanation which should be viewed as a supplement, rather than a rival, to the traditional approaches to explanation.

VI

As we have remarked (see note 3), there are many suggestions in van Fraassen's text that he does not intend to offer an "anything goes" account of explanation. In the last section, we have attempted to show that this intention ought to commit him to solving most (if not

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

all) of the traditional problems of the theory of explanation. We want to conclude by considering an obvious question. If we interpret van Fraassen as supposing that there are constraints on why-questions and their answers, how does this affect the general argumentative strategy of *The Scientific Image*?

Van Fraassen's discussion of scientific explanation is part of an effort to show that theoretical virtues beyond the saving of the phenomena are pragmatic. That argument eliminates a certain strategy for defending theoretical realism. If the realist proposes that (i) there is an objective criterion of explanatory power that distinguishes among empirically equivalent theories and (ii) theories with greater explanatory power have a stronger title to belief, then the doctrine of *The Scientific Image* appears to oppose the proposal by denying (i).[5] If we are correct in our assessment of van Fraassen's position, then it seems that the realist can get at least as far as (i). For, if it is once granted that we can produce statements which favor the topic of a why-question but which do not stand in any objective relevance relation to that topic (or, more exactly, to the ordered pair of topic and contrast class), then it appears that a theory may save the phenomena without generating answers to why-questions founded on genuine relevance relations.

We have argued that, if he is to avoid the "anything goes" theory of explanation, van Fraassen must offer a characterization of objective relevance relations which, in effect, overcomes the traditional problems of the theory of explanation. Now, within the traditional theories, there is ample room for prediction without explanations: we can have deductive arguments that fail to explain their conclusions, assemblages of statistical relevance relations that bestow high probability on a statement without explaining it. Once van Fraassen has introduced analogous distinctions within the theory of why-questions and their answers, through the provision of constraints on genuine relevance relations that separate mere favoring from the adducing of relevant information, there can be theories yielding statements that favor the set of topics in a given class (or even imply those topics) without generating answers to any genuine why-question with any of those topics. We would thus have the basis for claiming that such theories are objectively inferior to their rivals that do furnish explanatory answers.

[5] It is clear from a subsequent paper, in which he discusses Clark Glymour's views about explanation, that van Fraassen would also object to (ii). See "Glymour on Evidence and Explanation," in John Earman, ed., *Testing Scientific Theories* (Minneapolis: University of Minnesota Press), pp. 165–176.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The consequence would be that van Fraassen would have to revise his account of what it is to accept a scientific theory by adding the idea that acceptance involves believing that the theory has explanatory power as well as believing that it saves the phenomena (or, perhaps, believing that the theory offers the best tradeoff between saving the phenomena and having explanatory power)—indeed, he seems to take just this tack in the article cited in note 5. Since van Fraassen can still avail himself of a (different) distinction between acceptance and belief, this consequence should be seen as providing only the entering wedge for an argument for realism.

We conclude that, if van Fraassen avoids the Scylla of the "anything goes" theory of explanation, then he is plunged into what he would view as the Charybdis of supposing that there is an objective virtue of theories distinct from their salvation of the phenomena. From our perspective, Scylla is (to say the least) uninviting, but Charybdis feels like the beginning of the way home.

PHILIP KITCHER

University of California/San Diego

WESLEY SALMON

University of Pittsburgh

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS AND CRITICISM

MAINSPRINGS OF METAPHOR*

JOSEF STERN dismisses extensional theories of metaphor on the ground that substitution of coextensive terms does not always preserve metaphorical truth: "It may be a truism that the metaphorical depends on the literal, but this cannot mean that the extension of a term interpreted metaphorically simply depends on its extension interpreted literally."† True enough. But extant extensionalists are committed to no such thesis. To say that metaphor is determined extensionally is not to say that a term's metaphorical extension is determined solely or simply by *its* literal extension. Rather, the extensional references of associated literal and metaphorical expressions intertwine in fixing a term's metaphorical reference. Among the resources available to the extensionalist are the interpretations of related literal and metaphorical expressions,[1] secondary extensions,[2] mention selection,[3] exemplification,[4] and complex reference.[5] There are no recipes for determining metaphorical meaning. But there are heuristics that guide our search, providing cues and clues about which aspects of the context and background might be relevant.

Stern notes that substitutivity of literally coextensive terms fails to preserve metaphorical truth. Although

(a) Juliet is the sun.

is true,

(b) Juliet is the largest gaseous blob in the solar system.

* Catherine Elgin would like to thank the University of North Carolina Research Council for support of this project.

† "Metaphor as Demonstrative," this JOURNAL, LXXXII, 12 (December 1985): 677–710, pp. 683/4.

[1] Nelson Goodman, *Languages of Art* (Indianapolis: Hackett, 1976), pp. 71–80.

[2] Goodman, "On Likeness of Meaning," in *Problems and Projects* (Indianapolis: Hackett, 1972), pp. 221–230; "On Some Differences about Meaning," *ibid.*, pp. 231–238; *Languages of Art, op. cit.*, pp. 204/5; *Ways of Worldmaking* (Indianapolis: Hackett, 1978), p. 104.

[3] Scheffler, *Beyond the Letter* (London: Routledge & Kegan Paul, 1979), pp. 31–36 and p. 142, note 97; "Four Questions about Fiction," *Poetics*, XI (1982): 279–284.

[4] *Languages of Art, op. cit.*, pp. 50–58, 85–95.

[5] Goodman, "Routes of Reference," in *Of Mind and Other Matters* (Cambridge, Mass.: Harvard, 1984), p. 63ff.; see also, for further discussion of this and the previously mentioned devices, Elgin, *With Reference to Reference* (Indianapolis: Hackett, 1983), pp. 146–154 and elsewhere.

0022-362X/87/8406/0331$00.50 © 1987 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is false. Since the sun is the largest gaseous blob in the solar system, Stern concludes that extensionalism fails. He summarily dismisses secondary extensions, failing to appreciate the firepower they add to the extensionalist's arsenal.

A *secondary extension* of a term is the (primary) extension of a compound containing that term. The extension of 'sun-description' is thus a secondary extension of 'sun'. But the extension of 'sun-description' is not determined by that of 'sun'. Some sun-descriptions, such as those found in mythology, astrology, and ancient astronomy, are not true of the sun. Still, 'sun-description' has a determinate extension—a particular class of words and phrases. And, although we have no rule for its instantiation, sun-descriptions are readily recognized. Indeed, this state of semantic affairs is common: we have no rule for the instantiation of 'chair' either, but we recognize chairs without difficulty. To be sure, we probably cannot decide every case. We may be hard put to tell whether an odd linguistic construction is a sun-description and whether an odd material construction is a chair. But such difficulties do not impugn the determinacy of either extension.

Terms that agree in primary extension typically disagree in secondary extension. Although the primary extensions of 'sun' and 'largest gaseous blob in the solar system' are the same, their secondary extensions are not. 'Apollo's flaming chariot', for example, belongs to the secondary extension of 'sun', but not to that of 'largest gaseous blob in the solar system'. Literal meaning, Nelson Goodman suggests, is a matter of primary and secondary extension. So coextensive terms that differ in secondary extension differ in meaning.[6] It follows that, if metaphorical interpretation is a function of literal meaning, coextensive terms with different secondary extensions bear different metaphorical interpretations. Even if Romeo was an extensionalist, then, his assertion of (a) did not require him to accept (b). Extensionalists are committed to basing interpretation on nothing but extensions. But we are free to invoke any, and as many, extensions as we like.

A more perplexing problem is that the single term 'sun' itself bears disparate metaphorical interpretations. Juliet is characterized as the sun because she inspires passionate love; Achilles, because he is prey to awful fury. Metaphorical meaning, it seems, attaches to tokens, not types.

But how do literally coextensive replicas—that is, tokens of a single type—differ semantically? Mention selection provides the an-

[6] "On Likeness of Meaning," *op. cit.*, p. 227.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
swer. In a mention-selective application, an expression refers, not to what it denotes, but to mentions thereof. 'Unicorn' mention-selects unicorn-descriptions and unicorn-pictures; 'sun' mention-selects sun-descriptions and sun-pictures. But not every sun-description need be mention-selected by a given inscription of 'sun'. Tokens occurring in a work of Ptolemaic astronomy, for example, mention-select 'moving celestial body'; tokens in a work of Copernican astronomy mention-select 'motionless celestial body'. Literally coextensive replicas thus can, and often do, diverge in mention selection. Still, mention selection is extensional. There is a determinate class of expressions mention-selected by each token; and two tokens are co-mention-selective just in case they mention-select exactly the members of the same class.

Replicas that bear disparate metaphorical interpretations differ in mention selection. Inscriptions of 'sun' that apply metaphorically to Juliet mention-select expressions such as 'life-sustaining' and 'beauteous'; those which apply to Achilles mention-select expressions such as 'life-threatening' and 'terrifying'. The associations effected by the metaphorical application of an inscription are thus products of the literal mention-selective reference of that inscription. It is important to note that coextensive terms that differ in secondary extension also differ in mention selection. Unlike 'munchkin', 'wizard' mention-selects 'wise man'. So 'wizard' applies metaphorically to people to whom 'munchkin' does not.

Metaphors achieve their effects through likening. This involves yet another mode of reference—exemplification. If a symbol both refers to and instantiates a label, it exemplifies that label. And, if two symbols refer to and instantiate exactly the same labels, they are coexemplificational. For example, a paint sample that both refers to and instantiates 'vermilion' exemplifies 'vermilion'; and separate paint samples that refer to and exemplify exactly the same labels—say, 'vermilion', 'flat', and 'latex'—are coexemplificational. Exemplification, like the other modes of reference we have considered, is thus extensional.

Things that do not ordinarily function as symbols come to do so by serving as samples of, and thereby exemplifying, labels they instantiate. This is the key to metaphorical likening. In calling Juliet the sun, Romeo highlights features she shares with the (literal) sun. Through his characterization, he brings her to exemplify labels such as 'glorious' or 'peerless'. So, in the simplest case, a chain of reference links the literal and metaphorical extensions of a term via their joint exemplification of a label. Thus Sol, the literal referent of 'sun', and Juliet, the metaphorical referent of 'sun', are linked by their

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

joint exemplification of the label 'glorious'. Longer and more complex chains may also connect literal and metaphorical subjects. Juliet might exemplify a label that exemplifies a label that . . . is exemplified by the sun. And the labels exemplified may be literal or metaphorical. Any number of chains can be operative at once. A rich metaphor is inexhaustible in that additional chains of reference between its subjects may yet be forged.

Context influences the interpretation of metaphors in several ways. Normally, a term is applied as part of a scheme of implicit alternatives. And a single expression might belong to a number of schemes. 'Night', for example, can be opposed just to 'day', or to 'morning', 'afternoon', and 'evening'. Moreover, the extension of a token of 'night' varies slightly depending on which scheme is in play. Deep twilight belongs to the extension of 'night' under the first scheme, to the extension of 'evening' under the second. Settling the interpretation of a given token thus involves determining what kindred terms are, or might be, used in a given context. This is so whether the token functions literally or metaphorically. Rosalind is easily recognized as the referent of Romeo's 'moon', for the way has been paved by his calling Juliet the sun.

Interpretation often depends on precisely what words are used and how they are described. This is largely a contextual matter. Romeo's previous Juliet-descriptions and Rosalind-descriptions, along with his conduct *vis-à-vis* the two women, reveal that his love for Juliet has totally eclipsed his affection for Rosalind. So his metaphorical comparisons of the two may reasonably be expected to favor Juliet. We do not need his explicit avowal of (a) to recognize Juliet as the proper referent of 'sun', Rosalind as that of 'moon'.

Candidacy for exemplificational reference may also be circumscribed by contextual factors. The options for a correct interpretation of (a) are limited by the fact that it is uttered by a love-sick adolescent. We can expect the labels jointly exemplified by Juliet and the sun to be superlatives appropriate for describing objects of love and desire. They are unlikely to be predicates that belittle their referents. So, even though the sun is in fact a relatively insignificant star, contextual considerations exclude 'insignificant' as a contender for joint exemplification. Whatever Romeo is getting at in calling Juliet the sun, he is not conceding that, in the greater scheme of things, she is relatively unimportant. Such contextual considerations are plainly insufficient to determine exactly which labels are jointly exemplified. But, by restricting the candidate pool, they focus our search, directing our attention to a neighborhood in which a correct interpretation might be located.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Our primary objective in this paper has been to defend extensionalism against Stern's cavalier dismissal. The account we sketch turns out to be stronger than the one Stern proposes. Unlike Stern, we invoke no intensional entities; so our theoretical basis is more economical than his. And, if it accomplishes as much, the theory with the weaker basis is the more powerful. But, in fact, our theory explains more, and more that we particularly want to know, about metaphor. According to Stern, the most that semantics supplies for the interpretation of a metaphor is the advice: Look to context (697/8). It does not explain how a metaphor likens the literal and metaphorical referents of a term. This, for students of metaphor, is a (perhaps *the*) crucial question. And it surely seems to be a question about how metaphors function linguistically.

By recognizing that exemplification is a mode of reference, that words and other symbols are among the referents of our terms, and that secondary extension and mention selection depend on actual usage, not "lexical meaning," we can explain both metaphorical reference and metaphorical likening. Unlike Stern's theory, ours makes no use of a distinction between linguistic knowledge and collateral information. This strikes us as a good thing.

CATHERINE Z. ELGIN
ISRAEL SCHEFFLER

Harvard University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

The Thread of Life. RICHARD WOLLHEIM. Cambridge, Mass.: Harvard University Press, 1984. xv, 288 p. Cloth $20, paper $8.95.

In *The Thread of Life,* a revised and expanded version of the 1982 William James Lectures, Richard Wollheim provides accounts of imagination, memory, and phantasy along the lines of discussions in previous essays[1] and, appealing to these accounts, argues the inadequacy of a particular philosophical conception of the life of a person. Wollheim ascribes the conception he finds inadequate to *relational* theories of personal identity. On the conception in question, the nature or essence of the life of a person is to be specified by some (perhaps disjunctive) relation that holds between two events just in case they belong to the life of one and the same person. (One thing, for example, that might be true of two such events is that both are experiential memories of the same past event. So the relation "being experiential memories of the same past event" might enter, as a part, into an account of the life of persons on the conception described above.[2]) Put very briefly, Wollheim's argument against relational accounts of the life of persons is that such accounts fail to explain adequately the relation between a person and a person's life.

To show that relational accounts of the life of persons fail is the principal object of the discussions and arguments of the first chapter of *The Thread of Life.* The chapter begins with a discussion of the relation between being a person and being an animal of a particular kind. Wollheim argues that we accept an animal, an animate being, as a person if that being belongs to a species whose members are persons. The idea here is that a being is accepted as a person by virtue of some "structural" fact (8): a fact pertaining to typical members of its species and not just to the individual. Wollheim suggests that this fact is the way that typical members of the species lead their lives, and it becomes clear in subsequent chapters that what he means in speaking of the way persons lead their lives is what he elsewhere calls "the

[1] "The Bodily Ego," in Wollheim and James Hopkins, eds., *Philosophical Essays on Freud* (New York: Cambridge, 1982); "Imagination and Identification," in Wollheim, *On Art and the Mind* (Cambridge, Mass.: Harvard, 1974); and "The Mind and the Mind's Image of Itself," *International Journal of Psycho-Analysis,* L (1969): 209–220, reprinted in Wollheim, *On Art and the Mind.*

[2] Wollheim in his discussions of relational theory considers only "psychological relations" between events of a person's life—for example, the relation defined in the text—since only such relations, he believes, can be subjected to the test of our intuitions.

0022-362X/87/8406/0336$00.90 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

psychology of the person" (49), that is, the relationships among a person's mental states and mental dispositions of imagination, memory, phantasy, desire, belief, and so on. The structural fact that makes the members of a species *persons* is the fact that these beings have a particular psychology.

In his first chapter Wollheim distinguishes between a person as a member of a species and a person's life as a sequence of events. Relational theories propose accounts of the latter. More specifically, Wollheim presumes that a relational theory offers, in the relation or relations that it puts forward, an account of the nature or the essence of a person's life. I noted above that Wollheim attempts to argue a particular inadequacy in relational theories. In more detail, he argues, first, that a relational theory cannot provide a purely relational account of a person, that it must describe relation(s) between events each of which is *already* ascribed to a person. (Wollheim's arguments here would call into question a claim like Hume's in the *Treatise* that the notion of a person arises from the continuities of mental life. As it happens, I do not find Wollheim's specific arguments against what he calls "constructionist" relational theories convincing. But I want to consider other aspects of Wollheim's general argument against relational theory.) Wollheim argues, second, that a relational theory cannot of itself decide the relative importance of the two aspects of events it traces in the life of a person: on the one hand, the relation(s) of continuity between the events, and, on the other, the fact that the events are all ascribed to the same person. Only an account of how persons lead their lives, Wollheim argues—that is, only an account of the psychology of persons—could properly account for the presence of the two aspects traced in a relational theory. To explain adequately how a person (as Wollheim puts it, "the thing") is related to his or her life ("the result"), we must consider, Wollheim maintains, how the mental states and processes of a person's life—the mental states and the mental dispositions of imagination, memory, phantasy, and so on—bring about a life.

The above gives only the barest impression of what Wollheim considers the specific failings of relational theories. But it suggests something of what we might learn about such theories from his discussions of the psychology of persons in subsequent chapters of *The Thread of Life*. In particular, by providing a more adequate understanding of memory and of other mental phenomena, these discussions might give a more precise impression of the way that relational theories gain some plausibility from too simple a conception of the mind's relation to the past. For example, memory may figure as large as it does in many relational theories of personal

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

identity because it is thought to be a cognitive relation to the past and, apart from its failures, constant. However our feelings about a past event may change, our memory of it preserves something of the past: memory, unlike our feelings, provides continuity with our pasts. Wollheim's account of the psychology of persons would set against this simple picture of the continuity of memory an account of the way memory works to preserve the past. Wollheim argues that memory involves an element of feeling or affect—a response to the event remembered—and that this affective element accounts for many of the particular continuities and discontinuities of memory, for the persistence as well as for disturbances of memory. Dispositions of memory, like other dispositions, have histories that are not simply "histories of steady secular decay" (99). Wollheim's account of memory does not show the simple picture of memory to be *false*. Memory has a cognitive aspect, and, to the extent it is veridical, it provides a continuity with the past. But his account suggests that the simple picture of memory, in leaving out memory's affective element, misses what enables our memories to have the influence they have on our lives and so misses what *accounts* for the persistence of memory. A relational theory of personal identity which appeals to this simple picture of memory will fail to encompass something of the very nature and essence of our lives.

But if this is true, how might it matter? Wollheim, as I have noted, thinks that relational theories mean to provide an account of the nature or the essence of the life of a person. So viewed, a relational theory that appealed to the simple picture of memory would fail of its own purpose. But I do not think that relational theories do seek to provide accounts of the sort Wollheim presumes they seek. Wollheim remarks that relational theories of the life of a person are "metaphysical in inspiration" (11). But this is not yet to say what, precisely, prompts a philosophical question about personal identity, nor, consequently, what interest in the life of a person is expressed in such a question. Hume, for example, thought that, without the continuities and connections that our memory produces or discovers, we would not have our notion of a person or a self as a "continu'd object" (*Treatise*, I, 4, vi). This does not mean that Hume takes the continuities and connections produced or discovered by memory to be the nature or the essence of the life of a person. At most he takes them to be the nature or the essence of the identity, the continuity, that we ascribe to the life.

Wollheim, in arguing as he does against relational theories, does not, I think, address the particular interest in a person's life which is

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

expressed in Hume's and perhaps other relational theories of personal identity. He speaks of the relations proposed in relational theories as "unity-relations." Relational theories, however, as I am acquainted with them, are concerned not with something as broadly conceived as the unity of a life, but more specifically with personal identity (my conception of a sequence of events as mine, that is, as constituting a self) or with the continuity of a life. For this reason, Wollheim's arguments may miss their mark.[3] [Wollheim, I should note, also presents arguments of a sort different from those I am here discussing. He argues against the adequacy of the thought experiments employed in discussions of personal identity (chapter I), and he questions the intelligibility of the supposition in some of these thought experiments that memories can be transferred from one person to another (chapter IV).]

The account in *The Thread of Life* of the psychology of the person begins with a typology of mental states, mental dispositions, and mental activities (chapter II). Wollheim's principal concern in the chapters that immediately follow his introduction of this typology is to give an account in particular of the relation between mental dispositions and mental states and specifically of the relation between these in imagination (chapter III) and in memory and phantasy (chapters IV–V). In subsequent chapters Wollheim discusses a number of other topics that also pertain to the psychology of the person, among them: the understanding that persons have of their lives in leading them (chapter VI); the development in our psychology of "the moral sense" (chapter VII); the nature of our concern for our future lives (chapter VIII); and our attitude to death (chapter IX).

The mental dispositions that figure prominently in Wollheim's account (imagination, memory, phantasy) alter our behavior, beliefs, feelings, and so on, not directly—in the way, for example, that the dispositions underlying certain mechanical skills guide our actions—but by bringing about mental states that in turn alter our behavior, beliefs, etc. Such dispositions have what effects they have on us or in

[3] Even if Wollheim's general argument against relational theory misses its mark, his discussion of the psychology of the person may support a criticism of relational theory of another sort. One question we might have about relational theories of personal identity is this: How do they come in the first place to ask a question about personal *identity* or, more specifically, about the *continuity* of the life of a person? Wollheim's account of the psychology of the person may contribute something here. It suggests that we might ask: How important in the considerations that motivate a philosophical question about personal identity (continuity) is a notion of continuity already informed by a simple picture of memory and of mental life in general?

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

us by causing mental states that, in Wollheim's phrase, "further the role of" the dispositions. [The role of a disposition is "the way in which the disposition modifies the creature who has it" (49).] A mental state might further the role of a disposition in more than one way. It might contribute to bringing about whatever it is the role of the disposition to bring about (behavior, belief, etc.). But it might also further the role of the disposition by reinforcing the disposition. A feeling of anxiety, for example, might serve a fear by dissuading one from encountering what is feared; it might also reinforce the fear and its prohibition. Wollheim is interested primarily in the ways that the roles of dispositions are furthered by a particular sort of mental state: namely, iconic mental states, states that represent events. A number of claims he makes about such states in general and about the way such states relate us to the past are important:

(1) An iconic mental state, in representing an event, also has a tendency to produce "a residual condition," the condition ["cognitive, conative, affective" (72/3)] that I would have found myself in had I experienced ["lived through" (81)] the event.[4] My imagining, remembering, or phantasizing an event includes as a part something of the ways the event represented would affect me (something of the feelings and of the desires which I might have) if I experienced the event in actuality. Wollheim describes this fact about iconic mental states by appealing to an analogy with theater. Iconic mental states, he suggests, have features or aspects that correspond to the theatrical roles of director, actor, and audience. Specifically, the "residual condition" is to be likened to the feelings, desires, and so on, brought about in a member of an audience as he or she watches an action on stage.

(2) The way that iconic mental states further the role of the dispositions that cause them is a matter of the character of these states: specifically, it is a matter both of their content ("intentionality") and of what it is like to be in these states ("subjectivity"). These two characteristics together constitute the "phenomenology" of a mental state. Wollheim appeals to this notion of phenomenology in his discussions of iconic mental states, but it is for the most part the subjectivity of those states—more specifically, one aspect of that subjectivity, the residual condition—which is at issue in these discussions. For example, Wollheim discusses how the pleasure we might feel in imagining something, which is a residual condition of imagina-

[4] This is how Wollheim characterizes the residual condition in imagination. There are problems in carrying this characterization over to unconscious phantasies.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tion, can further the influence of desire. (Wollheim maintains that the dispositions that bring about states of imagination are desires.)

(3) Iconic mental states, in furthering the role of dispositions, may have effects on the course of the dispositions. Not only might they reinforce the dispositions; they might alter them. ["Outbursts of jealousy can strengthen jealousy: love satisfied can weaken love: hope can reinforce optimism: the exploration of phantasy can shake the hold of phantasy" (32).]

The three claims above suffice to show in general terms how Wollheim's conception of what it is to lead the life of a person differs from the conception in relational theory of a life as a sequence of events. On Wollheim's conception, the life of a person is held together not by a continuity of beliefs about his or her past, but by an interaction with the past. Dispositions are able to carry evidence of a past into the present because they carry the influence of the past into the present, and they are able to do that, in many cases, because they bring about mental states that leave us feeling something or wanting something. But dispositions that carry the influence of the past in these ways, through feeling or desire, are in turn susceptible to the influence of the present. [See (3) above.] Dispositions are a threading of our lives, not merely threads in them.

There are two further observations I would like to make about the three claims I cite above. The first concerns the sort of elaboration these claims receive in Wollheim's discussion of mental dispositions and iconic mental states. The second concerns an orientation that these claims reflect. To turn to the first matter: the three claims describe features of iconic mental states which Wollheim thinks can be found in each of the cases he considers, namely, imagination, memory, and phantasy. In making these claims Wollheim is summarizing features of examples which appear in various contexts of his discussion. He offers no general argument for the three claims, nor, perhaps, could he. What needs to be shown, above all, is how the particular role of each of the mental dispositions that give rise to iconic mental states is furthered by the phenomenology of such states. Since the different kinds of disposition, hence the different kinds of iconic mental state, have different roles, each kind needs to be considered in turn. I make this observation in part in order to suggest what evidence Wollheim's three claims might gain (and, to some extent, I think, do gain) from his examination of particular examples, but also in order to note that his discussion of these claims may fail adequately to address the simple conception of memory which enters some relational theories. Thus Wollheim fails to show

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that the second of his three claims applies to the specific dispositions and mental states of memory. The second claim, as I understand it, would maintain that the continuities and discontinuities of memory cannot be accounted for apart from accounting for the ways in which elements of feeling and of desire further the role of memory. Wollheim remarks that the affective element of memory "could account for whatever capacity event-memory has for uniting into a whole the different events that compose a life" (110). But he does not explain this. In particular, he does not explain how, precisely, the affective element of memories enters an account of the continuities and discontinuities, the recurrences and lapses, of memory.

The second observation I wish to make about Wollheim's three claims concerns, as I said, an orientation they reflect. Specifically, the emphasis in these claims on the residual condition (for example, the affective element of a memory) is an emphasis informed by psychoanalytic theory. At least, it reflects an emphasis that one might expect psychoanalytic theory to bring to an account of memory (remembering, forgetting), and it arises explicitly from psychoanalytic theory in Wollheim's discussion of the mental dispositions and states of phantasy. Wollheim's discussion of phantasy (chapters IV, V) includes accounts of projection and introjection, mental phenomena that are of central importance in psychoanalytic accounts of the way the child expresses feelings of love and hate in unconscious phantasy. Wollheim's interest in phantasy is twofold. In the first place, he is interested, generally, in the way phantasy preserves in the present the influence of past events (that is, the influence of past phantasies). His account of phantasy in this respect suggests another way in which relational theories appealing to a simple conception of memory fail to capture fully the manner in which the present of our lives is related to our pasts: phantasy, not only memory, preserves the influence of the past in the present. In the second place, Wollheim is interested in a specific feature of the mechanisms of projection and introjection, namely, the fact that in both mechanisms mental states and processes are represented as corporeal—more specifically, bodily—objects, products, or processes. Wollheim's account of what he calls, following Freud, "the thesis of 'the bodily ego' " (146)—the thesis that the mind represents, in unconscious phantasy, mental states and processes as bodily objects, products, and processes—suggests a quite specific way in which the mind, at least at early stages of its development, relates itself to the fact of its embodiment. What the thesis of the bodily ego begins to trace is how the mind—that is,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the infant—comes to terms with (if not fully to an understanding of) the fact of its embodiment.

Wollheim's account of what it is to lead the life of a person is an account of the psychology of the person: an account of mental states and processes such as those of imagination, memory, phantasy. Though informed by psychoanalytic theory, it is also informed and limited by what Wollheim considers to be the tasks of a philosophy of mind. His discussion of the psychology of the person takes its general terms ('state', 'disposition', 'intentionality', 'subjectivity') and, for the most part, its subjects (the nature of imagination, memory, and phantasy; the nature of our concern for the future; and so on) from philosophy. An exception of sorts is his discussion of "the growth of the moral sense" (chapter VII), in which he traces the origins of guilt and shame in the development of the superego. Wollheim prefaces this discussion with a claim that "the test" of whether someone's beliefs, sentiments, or decisions are a part of morality is whether they "appropriately descend from the relevant part of his psychology" (200). The claim here would oppose, at least on Wollheim's conception of the relevant psychology, a great deal more than the appeals of "mid-twentieth-century philosophy" to "the terminology that morality employs" (197). It also opposes traditional moral philosophies as divergent as Hume's and Kant's which would discover limits to the demands of sentiment, including the demands of the so-called "moral" sentiments (guilt, shame), in the capacity of reason.

One question we might ask about his own account of what it is to lead the life of a person is a question that Wollheim asks about other views of the nature of persons (views of persons as animals or views of their lives as successions of interrelated events), namely, What makes the process he describes the leading of the life of a *person?* What about the way we lead our lives makes this *our* way, the way that persons of the human species lead their lives? Wollheim does not offer a specific answer to this question; that is, he does not offer any beyond whatever answer is suggested by his choice of subjects to include in *The Thread of Life* and by the emphases in his discussion of these subjects. Wollheim, I think, considers the relations he describes between mental dispositions and mental states and, further, the particular importance of the phenomenology of our mental states in these relations to be not only central facts of the way we lead our lives, but perhaps *the* central facts. But these facts alone would not, I think, adequately answer the question: What makes the way we lead our lives the way of a human being, the way of a (species of)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

person? Indeed, I do not think that this question has an answer within the limits defined by the subjects and the emphases of *The Thread of Life*—within the limits of what Wollheim considers a philosophical account of the psychology of the person. I might describe what this account leaves out as certain subjects: broadly, the fact that our lives have characteristic paths, that they have characteristic crises and turnings (Wollheim mentions an aspect of this fact, but does not elaborate on it, under Freud's rubric of "phases of psycho-sexual development"), and, further, the fact that some of our relationships to others are, as I might put it, public, not just matters of our psychology (although I might equally describe this as the fact that we possess certain capacities not studied by Wollheim: for example, the capacity to reason and, in Kant's sense, the capacity of judgment). But I might also say, since Wollheim addresses something of each of these facts in his own way, that Wollheim's account of our psychology imposes a particular, a specific understanding on them and so leaves out aspects of these facts crucial to their definition of our lives.

NORTON BATKIN

Yale University

ANNOUNCEMENTS

The trustees and editors are pleased to report the appointment of Michael Kelly as managing editor of the *Journal of Philosophy* as of July 1, 1987. Dr. Kelly received his Ph.D. from Boston University in 1986 and was managing editor of the *Philosophical Forum* from 1979–1984. He succeeds Leigh S. Cauman, who has been managing editor since December of 1961.

The editors and trustees are pleased to welcome Leigh S. Cauman, retiring managing editor, to the Board of Trustees, and Gisela Striker, professor of philosophy at Columbia University, to the Editorial Board. Arthur C. Danto has agreed to become president of the Board of Trustees, and Barbara Gimbel, treasurer. Akeel Bilgrami and Charles Larmore, associate professors of philosophy at Columbia, will become consulting editors.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

The editors report with sorrow the death of W. H. Walsh, fellow emeritus of Merton College, Oxford, fellow of the British Academy and of the Royal Society of Edinburgh. Professor Walsh was fellow and tutor in philosophy at Merton from 1947 to 1960. He was professor of logic and metaphysics at Edinburgh University from 1960 until his retirement in 1979. He died on April 7, 1987, at the age of 72.

The editors regret to report the death of Bruce T. Riley, Professor Emeritus of Philosophy and Religion at Youngstown State University, on February 19, 1987. Born in 1908, Professor Riley earned degrees from Cornell College in Iowa and from the School of Religion at Boston University, and in 1940 received a Ph.D. from Boston University. From 1947 until his retirement in 1974, he taught philosophy and religion at Youngstown University, and served as chairman of the department for some years.

The editors regret to report the death of John R. Myhill, Professor of Mathematics at State University of New York/Buffalo, on February 15, 1987. Professor Myhill earned a B.A. from Cambridge University, England, and a Ph.D. from Harvard University. Before joining SUNY/Buffalo in 1966 as full professor, he held positions at the University of California/Berkeley and at Stanford University. His work included important contributions to mathematics, logic, philosophy, computer science, theoretical linguistics, and computer music. Professor Myhill was 63 years old at the time of his death.

The Federation Internationale des Societés de Philosophie announces a call for papers for an Extraordinary World Congress of Philosophy, to be held September 20–26, 1987, in Cordoba, Argentina. Papers in any field of philosophy are welcome. Papers should not exceed 1800 words; deadline for submissions is June 30, 1987. Further information may be obtained from Congreso Internacional Extraordinario de Filosofia, Pabellon Residencial, Estafeta Postal #32, Ciudad Universitaria, 5000 Cordoba, Argentina.

The editors of the Aristotelian Society and the Mind Association are pleased to announce a joint session, to be held July 10–13, 1987, in Cambridge. Speakers include Ned Block, Myles Burnyeat, Jeremy Butterfield, John Campbell, Peter Clark, David Cooper, John Cottingham, Elizabeth Fricker, Graeme Forbes, Mary Hesse, Mark Johnston, Gregory McCulloch, Richard Rorty, Peter Smith, and Jonathan Suzman. Inquiries may be addressed to: 1987 Joint Session, Philosophy Faculty, Sidgwick Avenue, Cambridge CB3 9DA, England.

In July and August of 1987, Ian Charles Tipton will be resident philosopher at Whitehall Museum, the historic colonial home of Bishop George Berkeley during his years in America. Mr. Tipton is Reader at the University College of Wales, and author of *Berkeley: The Philosophy of Immaterialism*. Whitehall will be open on weekend afternoons in June and September, and every day from 10 to 5 during July and August. It is located on Berkeley Avenue between Route 138 and Green End Avenue, in Middletown, Rhode Island.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Applications are invited for the Tennent Caledonian and Royal Bank of Scotland Research Fellowships at the University of St. Andrews, Centre for Philosophy and Public Affairs. Inquiries should be addressed to the Director, Dr. Gordon Graham. Applications, including a CV, a research proposal, and the names of two referees, should be submitted by October 1 of the academic year preceding that in which the fellowship is to be held to the Establishment Officer, University of St. Andrews, Fife, Scotland KY16 9AL.

The North American Society for Social Philosophy is pleased to announce a call for papers for a session on Ethical Issues and the Holocaust, to be held as part of the Fourth International Conference on Social Philosophy. Inquiries may be addressed to Sander Lee, Fourth NASSP International Conference, Philosophy, Parker Hall 27, Keene State College, 229 Main Street, Keene, N.H. 03431.

The Board of Officers of the APA is pleased to announce the ninth biennial Matchette Prize competition, made possible by the support of the Franklin J. Matchette Foundation. The prize is awarded to the author of a book of outstanding philosophical merit published since the closing date of the last competition. Authors must have been under the age of forty at the time of publication. The deadline for receipt of nominating letters and copies of nominated books is September 1, 1987. More information may be obtained from: Shirley Anderson, Administrative Services Coordinator, American Philosophical Association, University of Delaware, Newark, Del. 19716.

The American Society for Aesthetics invites nominations, including self-nominations, for the office of Secretary-Treasurer, which is a five-year position with a limit of two consecutive terms. The American Society for Aesthetics is a scholarly organization of about 800 members, which publishes *The Journal of Aesthetics and Art Criticism* and *The ASA Newsletter* and which holds an annual national meeting and three annual regional division meetings. Nominations and supporting materials may be sent to Anita Silvers, Chair, ASA Secretary-Treasurer Nominating Committee, Philosophy Department, San Francisco State University, 1600 Holloway Avenue, San Francisco, Calif. 94132.

The Society for the Philosophy of Sex and Love announces a call for papers to be presented at its meeting at the Eastern Division of the American Philosophical Association, December 1988. Papers are limited to critical discussions of *Sexual Desire: A Moral Philosophy of the Erotic,* by Roger Scruton (New York: Free Press, 1986). Deadline for submissions is April 1, 1988. Inquiries may be directed to Alan Soble, Philosophy Department, University of New Orleans, New Orleans, La. 70148.

Philosophy in Context announces a call for papers for Volume 17. Papers should examine the relationship between philosophical theories and practical applications. Contributors may wish to consider historical questions: for ex-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
ample, the influence of patterns of reasoning in science, morality, or law upon philosophical theories; cultural, social, political, or institutional criticism is also welcome. Deadline for submissions is August 15, 1987. More information may be obtained from Joseph DeMarco, Editor, *Philosophy in Context,* Department of Philosophy, Cleveland State University, East 24th Street and Euclid Avenue, Cleveland, Ohio 44115.

The editors welcome the formation of a new scholarly journal to be called the *Journal of Agricultural Ethics.* The journal will publish philosophical or scientific papers clearly related to all ethical issues confronting us as a result of agriculture and agricultural products. Manuscripts to be considered for publication should be sent to either of the co-editors: Frank Hurnik, Department of Animal and Poultry Science, and Hugh Lehman, Department of Philosophy, University of Guelph, Guelph, Ontario, Canada N1G 2W1.

The Social Philosophy and Policy Center, Bowling Green State University, announces the second in a series of three bicentennial conferences on the general theme of "Economic Rights and the Constitution." These conferences are a component of the National Endowment for the Humanities Special Initiative for the Bicentennial of the U.S. Constitution. The conference, "Liberty, Property, and Government: The Nation and The Constitution Grow Together," will be held October 22–24, 1987, at Bowling Green State University. Participants include: Richard Epstein (Law/Chicago), Tony Freyer (History and Law/Alabama), Harold Hyman (History/Rice), Alan R. Jones (History/Grinnell), Michael Kammen (History/Cornell), Paul Murphy (History/Minnesota), Mary Cornelia Porter (Political Science/Roosevelt), Glen O. Robinson (Law/Virginia), and Harry N. Scheiber (History and Law/California at Berkeley). Further information may be obtained from Kory Tilgner, Assistant Project Director, Social Philosophy & Policy Center, Bowling Green State University, Bowling Green, Ohio 43403-0188.

The Department of Philosophy of the University of Connecticut is pleased to announce the appointments of Diana Meyers (Montclair) as Associate Professor of Philosophy, and of Leonard Katz as Assistant Professor of Philosophy, effective September 1, 1987.

Applications are invited for the Andrew W. Mellon Faculty Fellowships in the Humanities at Harvard University. The one-year appointment will commence July 1988. Application is open to nontenured junior faculty with at least two years of postdoctoral teaching as college or university faculty in the humanities. Ph.D. prior to June 30, 1986 is required. Special consideration will be given to candidates who have not recently had access to the resources of a major research university. Besides an annual salary, the appointment carries limited teaching duties and an opportunity to develop scholarly research. Applications are due November 2, 1987; awards are announced February 1, 1988. Further inquiries may be addressed to: Richard M. Hunt, Program
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Director, Harvard University Mellon Faculty Fellowships, Cambridge, Mass. 02138.

The Department of Philosophy and the Geneseo College Committee for Celebration of the Bicentennial of the U.S. Constitution announce a call for papers for the Nineteenth Conference on Value Inquiry entitled: The Federalist Papers: An Explanation and Defense of Ideas and Issues in the U.S. Constitution, to be held April 8/9, 1988. Especially welcome are papers concerned with particular constitutional issues in *The Federalist Papers,* such as impeachment, executive prerogative, and congressional involvement in foreign policy; intellectual sources of *The Federalist Papers,* such as republicanism, the mixed regime, federalism, common law tradition, juridical review, and subsequent constitutional interpretation. Deadline for submissions is September 10, 1987. More information may be obtained from: Director, Nineteenth Conference on Value Inquiry, Philosophy Department, State University College, Geneseo, N.Y. 14456.

The Department of Philosophy and Religion at Hollins College is pleased to announce that Nickolas Pappas (Harvard) will join the department as Assistant Professor of Philosophy, beginning in 1987/88.

The Departments of Philosophy and Theology at Marquette University will sponsor a conference, entitled Kant's Philosophy of Religion Reconsidered, on November 6–8, 1987. Among the speakers will be: Sharon Anderson-Gold (Rensselaer Polytechnic Institute), Diana Axelsen (California Lutheran), Ronald Green (Dartmouth), Leslie Mulholland (Memorial University of Newfoundland), Philip Rossi (Marquette), Wilhelm Vossenkuhl (Bayreuth), Allen Wood (Cornell). Commentators include: Joseph Runzo (Chapman), Denis Savage (Marquette), John Treloar (Marquette), Alexander con Schoenborn (Missouri), Nicholas Wolterstorff (Calvin College, and Free University of Amsterdam). Further information may be obtained from either Michael Wreen, Department of Philosophy, or Philip Rossi, Department of Theology, Marquette University, Milwaukee, Wisc. 53233.

The Center for Critical Thinking and Moral Critique, with support from the California Humanities Project, will sponsor its Seventh Annual and Fifth International Conference of Critical Thinking and Educational Reform, to be held August 2–5, 1987, at Sonoma State University. Conference registrants may take part in a pre-conference workshop for secondary school and community college educators, July 31 to August 1, 1987. More information may be obtained from Joel Rudinow, Project Coordinator, The Constitution, Critical Thinking and Civic Literacy, Center for Critical Thinking and Moral Critique, Sonoma State University, Rohnert Park, Calif. 94928.

The Department of Philosophy of Skidmore College is pleased to announce that Maria T. Carl (Marquette) has been appointed Assistant Professor of Philosophy, effective July 1, 1987.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIV, NO. 7, JULY 1987

RED AND RIGHT

JOHN McDOWELL has argued that an analogy between secondary qualities and moral properties, together with a proper account of the former, can suggest an account of the latter which supports moral realism.[1] Antirealists too may accept the view that these kinds of property are importantly analogous; but, if so, they will normally assume that realism in regard to colors, for example, amounts to naive realism, a confusion of secondary with primary qualities. According to McDowell, antirealists about colors typically think that appearances are colored and that the colors of appearances must resemble properties of objects in themselves if colors are real properties. They accept some variant of John Locke's account of primary qualities and also agree with Locke that nothing in physical objects resembles phenomenal colors as they are perceived. This denial is supported by the lack of any appeal to such properties in our best explanations for the ways we perceive colors. Since we have no inductive reason to suppose that objects have properties that resemble phenomenal colors, antirealists (of this sort) conclude that colors are properties only of appearances themselves.

The inductive argument, however, seems entirely superfluous here. As Bishop Berkeley pointed out, the notion of an experience or appearance resembling a physical object borders on unintelligibility. The idea that objects could have intrinsic properties which resemble colors but which, in being intrinsic, make no essential reference to the ways we perceive them, is at best difficult to conceive. As Berkeley realized, this account, rather than call into question the reality of properties such as colors, renders questionable its notion of primary

[1] "Values and Secondary Qualities," in T. Honderich, ed., *Morality and Objectivity* (London: Routledge & Kegan Paul, 1985). Page references in text refer to this article.

0022-362X/87/8407/0349$01.40 © 1987 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

qualities and, with it, the supposed distinction between types of property. Unlike Berkeley, we may be wary of positing mental objects with intrinsic properties, accepting only talk of objects appearing in certain ways. If so, dismissing the notion of resemblance will leave us, at this point in the argument, not only without any reason for questioning the reality of secondary qualities, but without any distinction between them and primary qualities.

I

McDowell reintroduces the distinction within a thoroughgoing realist framework by accepting Locke's positive account of secondary qualities without the contrast involving resemblance. We are to understand secondary qualities in terms of objects' being such as to cause them to appear to us under certain conditions in certain ways (111). Such properties differ from primary qualities, since there is no appeal in the analysis of the latter to the ways they appear. But the former are in another sense equally real, in that being such as to appear in a certain way under certain conditions is independent of actually appearing that way or being believed to appear that way.

Colors and other secondary qualities are not simply identified here with the causes of the ways they appear. An identification of redness, for example, with properties of microphysical surface structures and ambient light would omit the phenomenal, sensuous nature of the quality as perceived. In addition, it would miss what is essential to being red, since the microphysical basis for red objects could change while the objects remained red, if the change were compensated for by a change in our visual systems. In order to avoid these mistakes, being red is understood here as being such as to appear that way, so that there is essential reference to the phenomenal property in the analysis. The account nevertheless preserves our realist intuition about the property as an object of perception as well. When we perceive colors, we seem to see properties of objects themselves. The property of being such as to appear a certain way is a real property of objects, albeit a relational one.

According to McDowell, this realism is only reinforced by appeal to inference to the best explanation as justifying ontological commitment. Appeal to a secondary quality does not enter into the explanation for the way it appears directly, for such explanations would be vacuous in themselves. To say that a tomato appears red because it is such as to appear red seems neither explanatory nor informative (although when we add that it is such as to appear red to normal perceivers, this may suggest that the present perceiver's perceptual response is accurate). A dispositional-relational analysis that essentially appeals to ways of appearing bars secondary qualities from any

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

direct causal explanations, since ways of appearing are not among the causal powers of objects. But McDowell points out that we cannot deny the existence of a secondary quality understood in this way while offering whatever explanations we do for the way it appears. Following David Wiggins, he views this as the true test for ontological commitment via induction of this sort (118).

Let us then consider an account along these lines of moral properties such as goodness or rightness. Certain disanalogies will emerge immediately. Clearly these properties will be analyzed in terms of relations not to perceptual appearances, but rather to evaluative responses in normal subjects. Second, we cannot understand moral properties simply in terms of objects' being such as to cause these responses in normal evaluating subjects. This would imply that normal evaluators cannot be wrong in their moral judgments in normal conditions (much as normal perceivers cannot err in their judgments of colors in normal conditions, according to the dispositional-relational analysis). But as individuals we do oppose the majority on moral issues far more often than we would oppose them in ordinary judgments of color (and not simply because of the relative importance of these judgments). Thus, we must either build more into the notion of normal conditions in the moral context, i.e., turn these into ideal conditions, or, as McDowell suggests, think of moral properties in terms of being such as to *merit* certain responses in subjects (118).

Despite this last disanalogy, despite an evaluative term's showing up in this analysis of moral-property concepts, we can still conceive such properties as real according to the same criteria we applied in the case of secondary qualities. An object's being such as to merit a particular evaluative response is independent of our actually evaluating it in that way or believing it worthy of that evaluation. This analysis of moral properties, like its counterpart for secondary qualities, seems true to the phenomenology of moral experience and to the realist intuition of encountering properties to which our judgments correspond through that experience. Whether the explanatory test is met in the case of moral properties is a separate question. Can we coherently deny the presence of such properties when explaining our evaluations and the actions and policies based on them? McDowell seems to think that we would lose much of the intelligibility and rationality of our responses if we denied real properties of the sort suggested (120). I shall not pursue this question at this point. I first want to raise the deeper question of the ultimate coherence of the analysis as it stands.

Another line of attack for the moral antirealist would be to press

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the disanalogies between supposed moral properties and secondary qualities. Simon Blackburn[2] has responded to McDowell in this way. He points out, for example, that moral properties are supposed to supervene on natural properties in much the way that secondary qualities supervene on primary. But, although at least the sketch of a causal explanation for this relation is available in the case of secondary qualities, Blackburn questions its availability in the moral case. Second, whereas perceivers need not be aware of the supervenience relation or appeal to it to justify their judgments of color, subjects unaware of the need to justify differences in moral judgments by appeal to differences in natural properties would be defective evaluators.

These two disanalogies taken together appear to weaken the case for moral realism based on the appeal to analogy with secondary qualities. They call into question the claim that moral properties remain distinct from natural properties of objects while supervening on them (in just the way that secondary qualities remain distinct from primary while supervening on them). But McDowell's latest account, outlined above, appears able to accommodate these disanalogies by deducing them from the main disanalogy between causing and meriting the respective subjective responses. The explanatory accounts to which the latter relations give rise are not exclusively causal, at least not directly so. This explains why an evaluator must be aware of the relation and not simply causally influenced by the purely objective properties. A full causal account of the influence of moral properties as so conceived would need to make intermediate appeal to the rationality and sensitivity of evaluators (assuming that the notion of meriting certain responses could be unpacked in terms of how rational and sensitive evaluators would respond).

II

To see whether such reductions would support moral realism, I want to return in greater detail to the analogy with secondary qualities. Rather than seek further disanalogies, I shall claim that an incoherence in the account of secondary qualities plagues the moral realist as well. The difficulty shows up only when the general definitional sketches offered above are rendered more specific. If I am right about this problem, then our original antirealist appears to be right that the analogy between moral properties and secondary qualities weakens rather than supports the case for moral realism. The Lockean analysis of secondary qualities assumes that, for each specific color, an object of that color will appear a certain way to normal observers under normal conditions. This holds not only for broad

[2] "Errors and the Phenomenology of Value," in Honderich, *op. cit.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

color concepts such as that of redness, but also for more specific shades. We also assume that every red object is some more specific shade of red. But, taken together, these assumptions turn out to be incompatible with plausible interpretations of the empirical data regarding color perception.

Given the vast number of discriminable shades and the many variables, both external and internal to perceivers, which affect apparent shades, it becomes overwhelmingly likely that normal perceivers will see slightly different specific shades on the same surfaces in normal conditions. The variables include not only mixtures of wavelengths reflected from particular uniformly colored surfaces (many different mixtures will be indistinguishable), but also light reflected from surrounding areas and from temporally prior distal stimuli (accentuating simultaneous and successive contrasts), and the adaptive states of the subjects' neurons as well as their prior perceptual expectations. With respect to these internal variables, viewing particular colored surfaces over time tends to reduce their perceived saturations and increase those of subsequently perceived complementary shades. For this reason, small patches appear brighter than large patches. Color constancy, which causes objects to vary less in apparent color than is predictable from their reflected light, might seem to reintroduce a closer correlation between perceived colors and physical properties of objects themselves. Instead, it introduces further variation, since it varies with such factors as surrounding objects and the age of the perceiving subjects.

Precise matching tests bear out the prediction that subjects have different matching functions, that they see different specific shades on the same surfaces in the same external physical conditions, even in extremely artificial and controlled conditions.[3] Furthermore, the abilities of normal subjects to discriminate colors differ considerably: no particular subjects are best over the entire range. It therefore seems that our everyday agreements in ascribing colors such as red to objects result from the breadth of such terms and concepts, from the fact that they cover a very large number of discriminable shades. If the less obvious facts are as described above, then we cannot specify in any nonarbitrary way a class of normal perceivers and normal conditions such that objects have those shades of color which appear to such subjects in such conditions. Any plausible way of doing so would have us ascribing incompatible shades to the same uniformly colored surfaces. Narrower specifications to remove these inconsistencies would be arbitrary. Thus, we cannot understand secondary

[3] See, for example, R. Burnham, R. Hanes, and C. J. Bartleson, *Color: A Guide to Basic Facts and Concepts* (New York: Wiley, 1963), pp. 56/7.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

qualities in terms of objects' being such as to appear certain ways to normal subjects under normal conditions. The empirical facts seem to drive us to a nonrealist position on colors, to the claim that colors qualify only the ways objects appear.

Disagreement in ethics is more readily apparent than disagreement in regard to colors. This is because the breadth of such terms as 'rightness' and 'goodness' adds to, rather than hides, disagreement in the application of narrower terms and concepts. There is less consensus here regarding which narrower concepts fall under the broader ones and regarding the application of the broader concepts to particular states of affairs. In the case of colors, disagreements in ascriptions of supposed real shades were problematic for a realist framework because, if genuine and unresolved, they would force us (barring skepticism) to ascribe incompatible properties to the same objects. Before turning to more radically relativist responses to this problem, we must investigate whether lack of consensus on correct ascriptions of goodness and rightness similarly plagues the moral realist. In my view there is a difference in this regard between 'good' and 'right'.

In the case of goodness, let us follow the clues to an account suggested above and say, as an approximation, that goods are those objects which would be desired for each person (given each person's distinct psychology) by a perfectly rational, knowledgeable, benevolent, and sensitive spectator, i.e., those objects which merit desire, approval, and efforts to attain them. Goodness is the common property of these objects. Other properties of objects in the set of goods will constitute a diverse cluster, although common needs will produce some common elements among the goods for different persons. The common property so specified counts as real, since it is independent from our beliefs about it and even to some degree from our actual desires; yet its connection with motivation is readily apparent.

It is clear that goods defined in this way can conflict. They may not be all jointly realizable. But they remain in themselves all good. Conflicts will obviously occur in conditions of scarcity. Less obviously, even in conditions of abundance, different conceptions of the good may be equally rational yet not mutually attainable. This is because such conceptions will include among their objects not only material goods, but ideals of the good life which may inhibit their mutual realization. As claimed above, such conflicts do not defeat the description of all these objects and ideals as good in themselves. What would be problematic are conflicts within the same agents' conceptions of the good life; but such conflicts are to be resolved by

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

rational choice as reflected in the desires of the ideal observer for each agent.

When we turn to an account of rightness, conflicts among the interests of different agents grow more problematic. The problem is that policies and practices judged to be right or just often constitute ways of resolving or simply settling these conflicts. If we attempt an interpretation of rightness along the lines of the realist interpretation of goodness suggested above, we encounter incompatible ways of resolving these conflicts which cannot be ruled out by non-question-begging conceptions of ideal observers (or of "meriting approval," the notion that appeal to ideal observers was intended to capture). One impartial, fully rational, and sensitive observer might aggregate interests in the utilitarian fashion, for example, while another might recognize certain individual interests as constraints on such aggregation. Or one might take loyalty to family or friends above that to country, and another the reverse.

In its implications for a realist account of the properties in question, disagreement here is more like that revealed in the perception of specific secondary qualities than like that involved in different conceptions of the good. In the case of goodness, a properly relativistic account could remain realist, since there it was a matter of ideal observers desiring different goods for different individuals, not a matter of assigning incompatible properties to the same individuals. Now, however, we have different observers approving of incompatible distributive schemes for the same groups, distributive schemes all seemingly meriting rational endorsement (at least none is eliminable by constraints of rationality alone). We cannot say that distributions approved of by any such observer are right, since, for example, an aggregative and a nonaggregative distributive scheme cannot both be right for the same group. At least they cannot both be right if they are incompatible and rightness is a real property in the sense so far suggested.

What blocks the realist interpretation offered so far is precisely the same problem that blocked it in the case of secondary qualities: the most plausible way of ascribing the property under this interpretation leaves us assigning incompatible qualities to the same objects. If we make the move to nonrealist interpretations, the details of the accounts may differ in the two areas. We might continue to ascribe color properties to ways of appearing and to hold beliefs about colors to be true when in correspondence with such appearances in certain conditions. We might, by contrast, think of the truth of judgments about right conduct in terms of coherence with a body of other settled judgments for groups of evaluators. But, in both cases,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

it seems that we must regard the phenomenology of the experience which generates judgments ascribing the properties to objects as mistaken, as a matter of projecting subjective perceptual or evaluative responses onto their objects.

III

There remains, however, another response for realists about colors and moral properties. They can accommodate disagreement by further relativizing the relations in terms of which these properties are analyzed. The analyses suggested above are already implicitly relative. Normal perceivers and ideal evaluators among humans, for example, may respond very differently from normal and ideal members of other species. That red objects for us are green for Martians would not trouble the realist about secondary qualities. More refined data on perception and reflection on moral judgment has revealed possible differences among responses within the class of normal human perceivers and ideal evaluators. We can avoid ascribing incompatible properties to the same objects in light of such differences by relativizing ascriptions to groups of perceivers and evaluators who agree among themselves. According to this revised analysis, when I say that an object is a specific shade of red, I mean that it is such as to appear that shade to perceivers who perceive as I do under normal conditions. When I say that a particular distributive scheme is right, I mean that it elicits approval among ideal evaluators who share my moral framework.

Once we relativize in this more radical way, seemingly incompatible ascriptions of relational properties are seen not to be genuinely inconsistent. Contradictions now occur not when we ascribe red and not-red to the same surfaces, or right and wrong to the same distributions for the same societies, but only when we ascribe red and not-red relative to the same reference class of perceivers, or right and wrong relative to ideal evaluators who share a moral framework. When we refer to nonrelational, primary qualities, disagreement implies error on the part of at least one party. But this, it may be said, is a mark precisely of the nonrelational, intrinsic, or objective property, not of the real. The truth of judgments ascribing such properties does not depend upon the truth of counterfactuals that refer to subjective responses of any sort. Realists, we have seen, may contrast such objective properties with relational, partly subjective properties while maintaining the reality of the latter. By narrowing the reference classes of subjects to which the analyses refer, they can avoid contradictory implications.

It may be objected that the concept of realism or criterion of the real now becomes unclear. We had implicitly defined real properties above as those independent if not from all subjective responses, at least from beliefs or theories about them. This epistemic or semantic

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

conception of realism (the truth of judgments about the real is not relative to beliefs or methods of verification) has become standard since it was introduced by Michael Dummett.[4] But relativizing moral judgments to evaluators who share moral frameworks or sets of judgments or beliefs violates this criterion for the reality of moral properties. Realists must now offer a different criterion and must show that appeal to it still marks a genuine distinction.

McDowell at least hints at an alternative when he writes of a property's "being there to be experienced, as opposed to being a mere figment of the subjective state that purports to be an experience of it" (114). The distinction marked by use of the term 'real' is now that between merely seeming to be at least in part external to the subject and really being so. For the relativist analyses of both secondary qualities and moral properties, this distinction is maintained in part by the fact that objects having these properties are still conceived as such as to cause certain responses in subjects only under normal or ideal conditions. It therefore remains possible for subjects to err in their ascriptions of these properties, for objects to seem to have them when they do not.

On the other hand, there is no possibility here of entire communities or reference classes of subjects under ideal conditions being mistaken in their ascriptions of these properties, and so the stronger sense of 'realism' which implies the ubiquitous possibility of error or skepticism has indeed been relinquished. If the reference class for my ascriptions of colors is limited to those who perceive always exactly as I do in daylight (a class that may consist only of myself), then the distinction between seeming to be red to me in daylight and actually being red will vanish, and with it will vanish almost all traces of the usual realist position in epistemology. There remains, therefore, a tension between the narrowing of the reference classes, something to which the facts about apparent disagreements may drive us, and the claim to preserve those convictions which underlie a realist epistemology and metaphysics.

This same tension reappears when we shift to the explanatory test for real properties. Recall that McDowell has weakened that criterion from requiring that appeal to such properties enter best explanations for subjective responses or observations, to requiring only that the denial of the existence of such properties be incoherent with those explanations we do offer. Recall also his claim that denying the reality of moral properties renders our evaluations less intelligible and justifiable. If we no longer believe that objects are such as to merit these responses, then we cannot explain their rationality, whatever mechanical explanations might be offered. Hence, we lose a

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

[4] "Realism," in *Truth and Other Enigmas* (Cambridge, Mass.: Harvard, 1978).

Digitized by Arya Samaj Foundation Chennai and eGangotri

degree of coherence in our understanding of ourselves. But is this claim plausible when we relativize in the radical way now envisaged? Although the fact that an ideal evaluator would approve of *x* might account for the rationality of my approval, the fact that *x* is such as to elicit approval in an evaluator just like me in moral outlook fails to account for any aspect of my approval. Although the presence of such a property in the object of my approval is undeniable, the point of affirming its presence or reality must remain obscure. Hence, we note again the tension between increasing relativity in the analysis of moral properties and the insistence that such properties nevertheless remain real.

Michael Smith, in a more recent defense of the dispositional-relational account of both secondary qualities and moral properties, defends the compatibility of realism and relativism.[5] He proposes yet another criterion for the reality of properties, one that avoids appeal to explanation altogether. According to him, if our experience is phenomenologically as of a property *P* of object *o*, then we have a defeasible reason to believe that *P* is a genuine property of *o*. Since our moral observations seem to encounter moral properties, we have reason to accept a realist moral metaphysics, and the relativist account coheres best with our other metaethical beliefs. But here again the degree of relativity is relevant to the claim of the account to remain realist. Can any apparently real property fail Smith's test if we narrow our reference classes sufficiently? Is there any possibility left for the error of projecting our subjective responses onto the world? Antirealists such as David Hume and J. L. Mackie agree that our moral experience is as of real properties; yet they believe it to be phenomenologically misleading. If real properties need be only such as to elicit those responses in us which they do, even those responses we make in our more dispassionate, informed, and reflective moments, then the error of projection becomes impossible. But then once more it becomes difficult to see what distinction the realist is drawing.

The point of asserting a realist position in regard to relational moral properties thus seems to depend on the size of the reference classes to which we relativize. Yet there seems to be no principled way to determine the extent of the relevant groups. Whether two subjects share the same moral framework and differ only in regard to particular judgments, or whether they differ in moral outlook or framework itself may not admit of an objective answer. We can determine whether, for the purpose of settling a particular seeming disagreement between them, we should regard them as involved in a

[5] "Can a Moral Realist Be a Moral Relativist?", read at Bertram Morris Colloquium on Social Philosophy, University of Colorado, 1986.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

genuine moral disagreement. To determine whether a particular dispute is genuinely moral, we ask whether the parties disagree in their moral evaluations even after they agree on the empirical facts. To determine whether the disagreement should be considered genuine or relativized away, we find out whether the parties are willing to reason from the set of judgments and principles on which they agree, to the solution most coherent with that body of data. If so, then for the purpose of that particular argument, they share the same moral framework. If not, then they do not. But relativizing this determination to particular moral disputes leaves even more puzzling the sense in which we can continue to characterize their assertions as referring to real moral properties. We have seen that the point of this characterization varies with the size of the reference classes to which objective correlates of the properties in question relate, and that the size of these classes itself is relative to the nature of particular disputes.

IV

For each subjective response to an object which we make, there is a correlative relational property of the object which consists in its being such as to elicit that response. Is there any advantage over an emotivist account in saying that our moral judgments refer to such properties? If the analysis refers only to responses under ideal conditions, then it can explain the rationality of some of our responses in real circumstances. But a suitably modified nonrealist or emotivist account might serve the same purpose. We can justify our responses, without mentioning properties in objects at all, by noting that they match those of an ideal observer.

McDowell and Smith argue that the realist accounts capture the phenomenology of perceptual and moral experience, whereas the nonrealist accounts do not. The nonrealist accounts attribute the error of projection to the naive perceiver and evaluator. But is the realist analysis really true to their experiences and so preferable on this ground? The phenomenology of perception seems to suggest not only that secondary and moral properties are real, but that they are nondispositional and nonrelational, in other words objective. The relational-dispositional account does not interpret secondary or moral properties as objective. Appeal to subjective appearances or responses does not simply fix reference to rigidly designated objective properties; it cannot be eliminated from the analyses. Thus, these analyses too fail to cohere with naive perceptual and moral consciousness and cannot claim a major advantage from that quarter. Red and right look to be intrinsic, objective properties; but neither antirealism nor the realist accounts under consideration here explain such seeming directly in terms of being so.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

It can be replied that common sense and phenomenology, while

Digitized by Arya Samaj Foundation Chennai and eGangotri

realist, do not subscribe to objectivist conceptions of secondary qualities or moral properties. The relational, as opposed to the nonrealist, account accuses naive consciousness of no error, if the latter presupposes no objective or nonrelational analysis. And, it can be argued, a property cannot look nonrelational, intrinsic, or objective unless the subject has those concepts as well as their opposites. Naive perceivers and evaluators are oblivious to metaphysical distinctions among types of properties. More generally, subjects use both perceptual and nonperceptual concepts without assuming either proper or improper analyses of them.[6] Knowers and evaluators of claims to knowledge, for example, do not err for failing to assume the correct analysis of knowledge, since they assume no incorrect account either. Thus, according to this response, the realist relational accounts of secondary qualities and moral properties are neither implicit in untutored perceptual consciousness nor out of tune with its phenomenology. But their nonrealist rivals *are* out of tune, since such consciousness remains implicitly realist, for secondary and moral properties look to be properties of objects.

In my view, this response imputes a convenient, but implausible and ad hoc, degree of conceptual sophistication to ordinary perceptual and moral consciousness by regarding them as realist but not objectivist in their suggested or assumed ontologies. Although subjects may not formulate distinctions between objective and subjective, nonrelational and relational properties, they do register surprise upon learning that colors are different sorts of properties from shapes. If there is a naive ontology to be captured, it represents these qualities as on a par. At least from childhood we know that what determines the shape of an object is not observer dependent; and we tend to assume the same for color, as we realize from introductory modern philosophy or history of science courses.

Any account that distinguishes primary from secondary qualities therefore departs from phenomenology and the naive ontology it suggests. Still, it might be claimed, the relational account remains closer to naive perceptual and moral consciousness than does a nonrealist account. Is this an advantage, once we learn that naive consciousness is in error with regard to these properties? No more so than a choice of spatial coordinates which came closer to capturing our intuition that the earth is flat could thereby claim an advantage over alternatives.

We must turn away from phenomenology to find a reason for preferring the relational account of moral properties over nonrealist interpretations of moral discourse. Stronger realist accounts might

[6] This response is made by Colin McGinn, *The Subjective View* (New York: Oxford, 1983), pp. 118, 134–136.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
appear to capture better the usual goal of such discourse, which is to secure agreement on morally controversial issues and, hence, convergence in behavior. According to an objectivist account, given two incompatible moral descriptions of a situation, only one can be true. This view may encourage disputants not to remain complacent in the face of disagreement. If we accept the relational account, by contrast, it is open to us to relativize away apparent disagreements, and this motivational advantage of the stronger realist view is lost. Conversely, on this view, discovery of real (relational) moral properties does not guarantee convergence of judgment or behavior. The goals of behavioral convergence and rational settling of disputes themselves may motivate us to view moral disagreement nonrelativistically, to reason from our shared convictions so as to share moral frameworks for particular issues. But such incentive is indifferent between relational and emotivist moral metaphysics. According to Smith, the fact that the teleology of practice gives us reason to resist accepting the implications of the relativist relational account affords no reason to believe it untrue (*op. cit.*, 17–18). But we still seek a reason to believe it as opposed to equally relativist nonrealist accounts.

Purposes of explanation afford no reason for preference either. Given knowledge of subjective response systems, we can infer to objective properties as part of the best explanations for the ways they appear. But we have noted that properties relative to particular subjective responses must be explanatorily idle. For purposes of explanation, we must infer to the objective correlates of the relational properties, to the causes of the responses. Appeal to responses of ideal subjects can explain the rationality of real responses, but we saw that this type of explanation suffers also when we are forced to relativize more narrowly.

Secondary qualities again provide parallels to moral properties. Were we to relativize radically our realist, relational conception of color properties, we would lose the typical point of color ascription —identifying their objects. When we must distinguish close determinate shades, even if we view these as (relationally) real, we can communicate the identities of their objects by appealing to them only if we know how they appear to other subjects. In the absence of such knowledge, it is better for purposes of identification to speak of purely objective properties, for example, reflectance properties of surfaces. In most everyday contexts, our broad color terms suffice for purposes of identification. But, once more, that sufficiency does not signal the reality of the properties to which we seemingly refer: we can identify objects by the ways they appear, when we agree in use of terms on the basis of those appearances.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar
Colin McGinn claims that the realist relational account of second-

Digitized by Arya Samaj Foundation Chennai and eGangotri

ary qualities explains why perception of them affords us knowledge of real objects, rather than only of our own experiences. According to him, the perception of secondary qualities represents the union of a perceptual system having certain intrinsic features with external objects. Because the qualities perceived are properties of real objects, a direct realist view of perception is tenable, and skeptical challenges to perceptual knowledge are not insuperable (*op. cit.*, 102/3). Once again, however, the nonrealist can respond that the realist relational account of secondary qualities is not needed either to explain perceptual knowledge or to mesh with a direct realist view of perception. We can justify claims to know the real properties of objects on the basis of the ways they appear if we can infer to the objective causes of their appearing those ways. (This is not to claim that we so infer unconsciously when we perceive, or that such justification is necessary for perceptual knowledge.) Perception of physical objects is direct, that is, unmediated by sense data, but it is also causally mediated. This causal relation *is* direct contact with objects in which they appear in certain ways. Not all ways they appear need match or enter into real properties of the objects in order for contact with them to be direct.

In sum, I can find no reason to accept a realist relational account of either secondary qualities or moral properties as opposed to nonrealist accounts, once we are forced by the facts of disagreements to relativize narrowly the truth of judgments in these domains. The important metaethical debate, as far as practical implications go, is neither a debate over meaning (as conceived earlier in this century), nor one over realism as defined in these last two sections. The distinction between objectivism and subjectivism is important in its import for moral understanding and practice. But here, given the connections between moral judgments, attitudes, and motivations (at least of ideal agents and observers), the prize seems too easily won by the subjectivist.

There remains, however, an intermediate ontological issue that is neither irrelevant nor without practical interest: realism as defined epistemically. The truth of moral judgments might be relative to subjective responses, but not to the moral beliefs of evaluators. This criterion for realism therefore remains the most interesting in the moral domain, as elsewhere. But, as I concluded at the end of section II, there appears to be more hope that it can be satisfied in the case of goodness than in the case of rightness.

ALAN H. GOLDMAN

University of Miami, Florida

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS AND CRITICISM

EPISTEMIC TIT FOR TAT

IN a recent paper, John Hardwig* has argued persuasively against epistemic individualism and for a more collective basis of knowledge, by analyzing appeals to expert opinion. He proposes two conclusions between which the reader is invited to choose, both of which are unpalatable to the epistemic individualist: the first is that one can know vicariously; the second is that only a community can know. In the present paper, I propose (I) to sum up some of the main points of Hardwig's argument; (II) to recall a number of points discovered in the course of investigations into the Prisoner's Dilemma; and (III) to suggest that the kind of trust that is needed in any such collective system of beliefs can be illuminated by these results stemming from the Prisoner's Dilemma, provided that the concept of trust be taken not in the moral, but in a strategic sense.

I

One of the standard definitions of knowledge requires that it be *true,* that it be *justified,* and that it be *believed.* On the face of it, this means that only the direct knower, in whatever field, can be properly said to *know* anything; for only the direct knower has direct evidence capable of justifying his or her belief. If I rely on expert *A*'s opinion on a subject in which I have no expertise, can I be said to know that fact or that theory? How about the expert himself, since he must take, on trust, a great number of empirical results obtained by others which he cites in his own papers? Hardwig shows that the lay person cannot be an epistemic individualist: he or she either has no background whatsoever in the area of expertise, has neither the time nor the inclination to master it, or simply does not possess the necessary intellectual acuity or dispositions to comprehend it. It is in most cases impossible for me, as a layperson, to do away with the trust I invest in the expert: I may be forever in debt to the expert for my beliefs and, thus, never directly *know* in these areas. Nonetheless, it is cognitive suicide to suppose that I do not know (even if indirectly) a great mass of data gleaned from the writings of experts.

This indirect knowledge, however, undermines epistemic individualism on one front; for, according to that doctrine, real knowledge must be guaranteed by my own justifications, not by someone else's. What guarantee have I that the other is reliable? I must be indepen-

* "Epistemic Dependence," this JOURNAL, LXXXII, 7 (July 1985): 335–349.

0022-362X/87/8407/0363$01.50 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

dently able to verify any results obtained by another, at least in principle. Hardwig has shown that, even if this is possible in some cases, even if it were possible in the vast majority of cases, there is still *some* area where I must defer to someone else's considered opinion, or simply admit that I do not know, cannot know, and therefore cannot base any action on my part on this indirect evidence provided by an expert opinion I cannot myself corroborate.

This also undermines epistemic individualism on another front; for even the expert in his or her own field relies on results obtained by other experts which are necessary to his or her own work. No expert can independently redo all the work done by others that constitutes the base upon which that particular expert builds. So science itself is not immune to this attack on epistemic individualism. Whether for the layperson or for the expert, the phrase '*A* has good reasons to believe that *B* has good reasons to believe that *p*' is not eliminable, in all cases, in favor of '*A* has good, directly obtained reasons to believe that *p*'.

Thus, rational inquiry must include at least some cases of deferring to authority. But this is tantamount to admitting that, under the traditional individualistic analysis of knowledge, it is sometimes rational to be irrational; for, if I believe something without proper grounds, I believe irrationally. Although this conclusion may be found minimally acceptable in the case of the layperson who perhaps cannot be expected to have all the relevant information that the expert is supposed to have, this form of cognitive elitism cannot avoid the fact that *no* expert has verified each and every piece of data upon which he bases his or her own expert opinion. So, suggests Hardwig, either no one knows much of anything (certainly not any scientific thing), or there is genuine knowledge reposing not within the individual but within the collective body of the community.

> We must recast our conception of what it means for beliefs and persons to be rational. We must also either agree that one can know without possessing the supporting evidence or accept the idea that there is knowledge that is known by the community, not by any individual knower (349).

II

The Prisoner's Dilemma was concocted in the 1950s by Melvin Dresher and Merrill Flood,[1] and has since been the subject of numer-

[1] I follow Robert Axelrod's exposition in *The Evolution of Cooperation* (New York: Basic Books, 1984). Interested readers may consult the excellent summary by Douglas R. Hofstadter, "The Prisoner's Dilemma, Computer Tournaments and the Evolution of Cooperation," *Scientific American*, CCXLVIII, 5 (May 1983): 16–26;

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ous studies. One of the more interesting for the purposes of this paper is Robert Axelrod's tournament,[2] to which we shall presently return, pitting various computer programs against each other in order to discover which set of tactics is the most advantageous in an iterated Prisoner's Dilemma.

Douglas Hofstadter has given a very intuitive description of the Prisoner's Dilemma, and I can do no better than to report it here. Suppose you have arranged to deal with a supplier in a rather strange way. Instead of meeting face to face with him and openly exchanging your money, which he desires, for his goods, which you desire, it has been agreed that you will leave the money in a bag at a prearranged spot in the woods somewhere, and that your supplier will leave the goods in a bag at another spot. After the initial agreement, you will never meet again. You have agreed to transact in this fashion every month. Should you keep your part of the bargain? Should your supplier?

On the one hand, you might argue, if you leave the money (cooperate) while he leaves an empty bag (defects), then you have lost; so it is better that you defect. On the other hand, if he cooperates, then you would do better to defect, since then you would have both his goods and your money; so, again, it is better that you defect. But your supplier can also reason in exactly the same way. The net result of the impeccable logic used by both parties is that both leave empty bags, and therefore leave the woods with nothing more than that with which they began. But, if both had cooperated, both would have gained. In other words, each can argue that, if the other party cooperates, it is better to defect in order to maximize gain, and, if the other defects, it is foolish to cooperate. But, then, no one cooperates and no one gains.

In order to clarify the situation for each player in this game, it is customary to construct a payoff matrix (see Figure 1) wherein each player's gains and losses are tabulated for both possibilities, cooperation and defection.

---

reprinted as ch. 29 of Hofstadter, *Metamagical Themas: Questing for the Essence of Mind and Pattern* (New York: Basic Books, 1985); further references will be to the book. For an excellent review article of Axelrod's book, see James L. Gould and Carol G. Gould, "Tit for Tat," *The Sciences,* XXIV, 6 (November/December 1984): 47–52.

[2] Reported in Axelrod, *op. cit.* See also Axelrod and William D. Hamilton, "The Evolution of Cooperation," *Science,* CCXII, 4489 (March 27, 1981): 1390–1396. Other accessible texts on the Prisoner's Dilemma are: Morton Davis, *Game Theory* (New York: Basic Books, 1970, 1983); and Anatol Rapoport, *Two-person Game Theory* (Ann Arbor: Michigan UP, 1966).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

		Player B	
		Cooperates	Defects
Player A	Cooperates	(3, 3)	(0, 5)
	Defects	(5, 0)	(1, 1)

Figure 1

If both players cooperate, then each gains by the transaction. If you are Player *A*, then you have left the money in your bag, and the supplier, Player *B*, has left the goods in his bag. Let us arbitrarily value this gain as 3 points for each of you. If you cooperate, but your supplier defects by leaving an empty bag, then he gains your money (5 points), while you gain nothing (0 points). The situation is reversed if you defect but he cooperates. If you both defect, then you each are awarded 1 point, perhaps as an indication of the admittedly dubious gain of having cheated the opponent while coming out no poorer than before. The actual values entered as payoffs in the matrix do not matter, as long as two conditions are met. If (following Axelrod) 3 points are allotted as *R*eward for mutual cooperation, 1 point as *P*unishment for both defecting, 5 points for the *T*emptation of defecting when the other cooperates, and 0 points as the *S*ucker's payoff for cooperating when the other defects, then the two conditions that make a payoff matrix a Prisoner's Dilemma may be succinctly stated:

(1) $T > R > P > S$
(2) $[(T + S)/2] < R$

If the matrix is redone to reflect only formal payoffs, then we obtain Figure 2:

		Player B	
		Cooperates	Defects
Player A	Cooperates	(R, R)	(S, T)
	Defects	(T, S)	(P, P)

Figure 2

Condition (1) ensures that the argument to the effect that it is better to defect no matter what the opponent does works for both, and condition (2) ensures that it is better to cooperate than to alternate defection and frustrated cooperation.

Now, if the agreement is for only one month, then there is no way of deciding which is the best strategy, if neither player can know what

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the other has done before leaving his or her bags. What about the reiterated case, where you both will be coming back month after month? The temptation to defect is always present, so your supplier may defect one month. What should you do the following month? You have already got the sucker's payoff once; should you give him the same treatment the next time? If so, will the mutual trust be forever shattered, leading to an absurd ritual of empty bags left in the woods month after month? Or will you both get stuck in a perpetual alternation between triumphant defection and frustrated cooperation? Or, finally, should you retaliate once or twice by defecting yourself, then attempt cooperation thereafter in order to show the supplier that you are not to be trifled with, but that you are willing to continue the relationship if he straightens out?

Game theory shows that there is no universally best strategy in this type of non-zero-sum game. No single strategy will guarantee a consistently favorable outcome in a reiterated Prisoner's Dilemma, when pitted against any and all possible strategies. That is why Axelrod decided to see what would happen if many different strategies, implemented as computer programs, were to compete against one another in reiterated Prisoner's Dilemma situations.[3] What is interesting for the purposes of this paper is the particularly simple strategy that turned out to be the best over-all performer in many such round-robin tournaments. This program, named TIT FOR TAT by its creator, Anatol Rapoport, can be succinctly described as *nice, provocable, forgiving,* and *clear.* It thus has all the earmarks of a *robust* strategy, that is, one that has survived many iterations and many generations in the artificial "universe" populated by the dozens of software denizens vying for supremacy.

TIT FOR TAT is not *the* best strategy, for there is no such thing. It is, however, a particularly resilient and successful one, given the varied environment provided by the other "players." First of all, TIT FOR TAT is *nice* because it always cooperates on the first encounter. However, TIT FOR TAT is no pushover, for it is *provocable:* it will retaliate by defecting on the next round if the other player has defected on the previous one. TIT FOR TAT is also *forgiving:* having once punished the offending player by defecting in turn, it goes immediately back to cooperating, and will continue to do so until the next defection when it will again retaliate only once. Thus, it does not hold a grudge longer than one move. Finally, TIT FOR TAT is *clear:* its strategy is easily learned and thus easily predicted by the other players, because its behavior is neither abstruse nor hidden behind

[3] The interested reader is invited to consult Axelrod's own presentations or Hofstadter's summary for more details on the tournament.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

complex rules, and it does not employ random generation of moves. All this behavior can be coded quite simply: Cooperate on the first move; thereafter, do whatever the other player did the previous move. The only "cognitive abilities" TIT FOR TAT has are the capacity to recognize previous partners, coupled with a memory of what transpired last time with that partner. As Hofstadter (729) aptly notes, TIT FOR TAT has no need for reflective rationality; purely reflexive reaction suffices.

What makes TIT FOR TAT a robust strategy is not that it won against the other players. On the contrary, it never *won;* at best it broke even or even lost a bit on each encounter. It proved more robust than the other entrants because it lost less, not because it won more. TIT FOR TAT is successful because it elicits cooperative behavior from the other player. It employs "friendly persuasion," or reciprocity, rather than crude egotism. Of course, there is no guarantee that the other player will catch on to TIT FOR TAT's strategy, or that it will see that it has more to gain with TIT FOR TAT by cooperating rather than defecting. There is no universal strategy in non-zero-sum games like the Prisoner's Dilemma. But TIT FOR TAT is a good strategy to gamble on, because it seems to have a rather good survival rate. The basic requirement is that the relationship be durable, that it not be a one-shot affair, or, as Axelrod says, that the *shadow of the future* be sufficiently long. A cooperative strategy like TIT FOR TAT has little chance in the short or medium term; only long-term relationships permit it to hold its own.

III

Hardwig left us with a difficult choice: either no knowledge is possible, or the only possible knowledge belongs necessarily to the community. If we are agreed that knowledge, albeit fallible, is possible, then individual knowledge cannot suffice. If knowledge by the community is to be the only possible kind, then only parts of the knowledge base will be known to individuals: no one individual can know everything. Worse still, the facts that even the expert individual knows depend upon things he or she does not know but accepts on trust. These are the things that are believed not because of direct evidence, but on indirect evidence, and they have the form: "*A* has grounds to believe that *B* has grounds to believe that *p*."

Reliance on another's beliefs seems to be a *sine qua non* for science. But why should I trust another's evidential report? Should not the required spirit be *not* to trust hearsay evidence? In an ideal world, one in which Plato would perhaps have been comfortable, a case could be made for the acceptance of facts only after having verified them oneself. But our real world seems to leave no room for doubting that the ideal world we may dream about is not ours to live

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

in. We must trust the evidential reports of others, simply because physically we cannot start from scratch. Whatever worth science may have, it requires this trust; for it is by and large a cumulative enterprise and no one individual can shoulder the evidential load.

Does this basic trust make science an irrational enterprise? Not at all. It does mean that epistemic individualism (or epistemic independence) is not the right model for giving a correct account of how cumulative science proceeds. Such an account is not, however, the purpose of this paper. Rather, I should like to pursue the question of trust and how it works in science and, indeed, in any cumulative enterprise where no one individual can claim to possess the totality of the evidence.

TIT FOR TAT is admittedly an idealized model of cooperative or noncooperative behavior. It neglects many factors important in human decision making, and, as such, cannot be taken very seriously as an exact description of how and why humans cooperate or defect, whether in day-to-day affairs or in the pursuit of science. If a human did conform to TIT FOR TAT's rules, then another player could predict exactly what that TIT FOR TATing human would do for any situation, providing information on the previous move was available. Unfortunately (or, fortunately?), no individual human seems to function precisely like pure TIT FOR TAT, if only because each and every other human does not. I may be willing to play a very robust strategy (in the sense explained above in the second section: nice, provocable, forgiving, and clear); but, if others are not persuaded to do likewise, I will certainly lose considerably and often. So, notwithstanding my good intentions, I will more probably diverge from the ideal model and bluff a bit.

My bluffing may take the form of being a bit less clear than I could be, in order to keep my opponents guessing and to maintain an advantage. I may be a bit more provocable than TIT FOR TAT would recommend, and punish defection more than once, or even cease to cooperate with a player if he or she defects too often and, thus, be totally unforgiving. Finally, I may decide not to be nice on a particular first encounter, even if it should be to my advantage to be so, simply because I do not like that particular player. So, although the name 'TIT FOR TAT' was coined in a human context, it seems that only computer programs really live up to its definition in the context of game theory. Humans seem to adhere to TIT FOR TAT much more faithfully in the negative sense of the expression, where retaliation is called for, than in the positive sense, where cooperation would be to the mutual advantage of both parties.

But is this really so? Is this rather pessimistic appraisal of humanity's refusal to espouse the virtues of cooperation accurate? From the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

perspective of the dog-eat-dog world of the business marketplace, where competition is fierce and corporate takeovers rife, a case could be made for such a noncooperative outlook. But it certainly is not the rule, nor could it be: members of the business world enter into verbal and formal contracts daily; it is assumed, until proved otherwise, that the customer will pay the bill on the date due; checks are accepted with a minimum of verification; currency is supposed genuine and not counterfeit, unless the authorities have alerted us to the existence of bad bills in the area. Of course, there is defection, and cooperation does not always follow the simplistic TIT FOR TAT format. But, by and large, cooperation is the norm rather than the exception, even in the supposedly tough marketplace. Although there are any number of ways to make a person part with his or her money, still, basic honesty is well regarded in the business world, and defection and dishonesty are mainly disapproved of.

In science, cooperation is of the essence. This type of cooperation does not require trust in the moral sense. It is not necessary to assume that trustworthiness is a moral virtue of the trustee, or that trust be construed as a "confident expectation of something; hope" (*Oxford English Dictionary*). This does not mean that moral virtues are to be excluded from science, or that science could have developed as it did without the moral rectitude of its participants. I should like to model trust as a *strategy,* rather than as some state of the truster perhaps related to some other state of the trustee. I certainly do not wish to deny the existence of any of these moral states. The idea is to see how far we can go in assuming that cooperation in the knowledge game is justified, even if the players of the game have no such moral virtues.

Since I cannot hope to justify all my beliefs personally, I must trust some experts. In a more traditional vein, I should inquire why I think I have reason to trust these people. In what follows, I hope to show that only cooperation, as defined above in game theory and as illustrated in the Prisoner's Dilemma, is necessary for the justification of vicarious knowledge. As Axelrod (182) has noted, faith in the future behavior of another player is less necessary than durability of the relationship: as long as the "shadow of the future" is sufficiently long, each player has less incentive to defect in order to maximize payoffs in the short run; for repeated encounters favor cooperative behavior simply from an egotistic viewpoint.

If an expert cooperates with the community, then I am justified in believing his or her reports within his or her field of expertise. Does this mean I should seek guarantees of his or her cooperation? Cooperation in science does not mean only working together; it means not defecting in the knowledge game. Defection in this context means to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

cheat by fudging, fabricating, or otherwise publishing unreliable results. Defection means succumbing to the temptation of leaving the other players in the knowledge game with the sucker's payoff, while attempting to maximize immediate gain. It means aiming for the payoff (T, S) to one's own advantage. The immediate gain for a scientific defector is the recognition that accrues to anyone who publishes in respectable journals. The temptation can be great; for, without publications, good jobs are scarce, academic tenure is impossible, and grants are not forthcoming. But, if an individual researcher cheats, the whole community stands to lose by this defection; for the reliance on data produced by other researchers is essential for the very existence of science.

If the individual researcher cooperates with the other members of the community, then the rules of the game imply that great care must be taken to ensure the exactness of the results. Infallibility is not necessary (nor is it attainable) as long as the members are ready to admit partial revisions in the light of further research; publishing what is true to the best of one's knowledge is sufficient. The payoff (R, R) is slightly less than (T, S) for the individual, but only in the short run. In the long run, not only does the community gain, but the individual gains also, for the same reasons that TIT FOR TAT did well in Axelrod's computer tournament. Cooperation breeds cooperation, to the mutual benefit of all concerned. Defection, on the other hand, though tempting in the short run, is not a good long-term strategy; for the community as a whole has a very effective strategy for dealing with defectors in the knowledge game. Reputation is very important in any social activity, but power and money can counterbalance a dishonest reputation to a certain extent. It is possible to continue to be successful in the business game of corporate raiding, for instance, even if the majority of business people condemn such doubtful means of turning a dollar; the money and the power such raiding brings, though they may not be everything, are the predominant factors in that game.

In the knowledge game, although money and power are not negligible aspects of what stimulates its participants to greater heights, they are not considered to be the fundamental criteria of success. Rather, what count are factual results that are reproducible. If results cannot be reproduced, they may simply be rejected: the individual researcher never gains any reputation at all. Worse, if it can be shown that the results have been willfully misrepresented, the punishment that ensues to the defecting researcher is lifelong exclusion from any gainful employment in the scientific domain. If money is the name of the game in business, cooperation is its name in science. Of course, it is possible to fool all the people some of the time, and

CC-0. Prof. Satish Domadia. Gurukul Kangri Collection, Haridwar.

Digitized by Arya Samaj Foundation Chennai and eGangotri

some of the people all of the time; but peer review and blind refereeing ensure that, in the long run, defectors should be found out. One must not underestimate the part played by aspiring researchers who have much to gain by punching holes in their elders' reputations. These latter reputations are usually difficult to undo, precisely because they have been built within the rules of the game. So scientific refutations, and scientific revolutions that depend partly on them, are not rife. But they *are* possible, since everyone serves as potential watchdog for everyone else.

What could traditionally be called trustworthiness or honesty is thus reinforced by the very real menace of exclusion. During apprenticeship, one soon learns that short-term gains of defection are far outweighed by the long-term rewards of cooperation, whether one is morally trustworthy or not. To be allowed to stay in the game requires cooperation, but the punishment is permanent exclusion. The reward for cooperation between the individual and the community (or the other members of the community) is not identical for the two parties. For the individual, the reward is reputation and all that can be obtained within society by one whose excellent reputation is deserved. For the fellow researcher, the reward is the sharing of data and theory which permits individual advances based upon cooperators' shared results. For the scientific community as a whole, the benefits are a growing body of knowledge with built-in corrective measures. For society at large, the benefits are the general reliability of results which are practically useful in all phases of life where greater and better information are always welcome. It does not pay, in the long run, to withhold data for too long, because others may publish their similar results first, thereby short-circuiting the desire to obtain priority. So it does not pay to be envious of the other players; for their gain is, in a very real sense, also the individual's gain: everyone's advances require the maximum input of reliable experimental and theoretical results. In a non-zero-sum game, it does not pay to try to maximize one's own gain by minimizing that of the other players, for doing so only ensures that long-term gains will be lower than cooperation would ensure. To win in science, it is not necessary to bring about someone else's loss: science is not a zero-sum game like chess.

What sort of strategy should I as a lay person adopt when I require information that I am not in a position to obtain or to verify on my own? It certainly would not be impeccably rational simply to follow the advice of the first expert upon whose opinion I happen to fall; for everyone is agreed that there are experts—and then there are experts. But, even if I am not an expert, it is not particularly difficult to ascertain whether the majority of experts are agreed on the partic-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ular point in which I am interested.[4] If there is a good consensus on the question, then it certainly is rational for me to follow their collective advice or to commit myself to a course of action based upon their concurrent opinion. Seeing that I am not an expert in the field, and that I am not in a position to gather the pertinent information I require for my practical needs, it would be irrational *not* to avail myself of the best knowledge base that humanity has amassed on the question, even if I am convinced that science, like all human endeavors, is fallible and subject to revision. The absence of absolute guarantees does not imply that randomized action is the only alternative.

In game-theoretic language, my following the experts' collective advice risks, of course, the sucker's payoff: maybe they are all wrong, and I will have made a bad move. My defection, that is, my refusal to heed this collective advice, would presumably hurt only me; but, if my action has repercussions on other people, then they also risk harm because of my bad gamble. My cooperation with the scientific community (supposing, in this case, that its advice is reliable) would similarly benefit me and others dependent upon my ultimate course of action; but it would not benefit the scientific community itself, at least not directly. Thus, my defection or cooperation do not seem to fit into even an informal Prisoner's Dilemma model, for my gains or losses seem to be totally independent of those of the scientific community. It may be true that, compared with the community as a whole, my personal case carries very little weight. But the same is true for each individual expert, except perhaps for the world-class leaders. What does count is the sum of the individual cooperative efforts, whether from the lay standpoint or that of the expert. It may not make much difference if I ignore expert opinion; but if a majority of lay persons ignored it, surely the social and economic fabric would reflect this mass defection. One result might be massive cutbacks in scientific funding, justified if no one believed in the enterprise any more. So, when summed with other individual actions, my small defection or cooperation does weigh in the balance.

The scientific community does not stand to gain at all from long-term defection with the public; for loss of credibility means loss of funding. Science cannot afford to permit wholesale defection within its ranks; so effective policing is necessary. Cooperation with society at large requires that the expert community spend the time and resources necessary to inform correctly, even if it is tempting not to

[4] See, for instance, Keith Lehrer, "Social Information," *The Monist*, LX, 4 (October 1977): 473–487, for a theoretical modeling of rational aggregation of social information. Even without such a formal apparatus, I can, so to speak, determine which way the collective wind blows to a degree sufficient for my immediate needs.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

do so for all sorts of reasons, not the least of which is the time necessary to popularize often abstruse results. Mutual cooperation requires that the experts' payoff, as well as the community's, be a bit less than it would be if the sole motivation were immediate maximization. The community's sacrifice takes the form of funding for pure research in domains where no one knows what benefits, if any, are forthcoming. The community's temptation would surely be not to stop funding but perhaps to limit it to applied research. But, if both partners are willing to settle for a bit less immediate gain (or a bit more immediate cost, which is the same), then mutual cooperation should ensure long-term gains satisfactory to both.

But what if the community of experts is not agreed on the particular point I wish to elucidate? What if there is no consensus, yet I need to decide on an important question that I should very much like to illuminate with expert testimony, but none is forthcoming, or, worse still, too much contradictory testimony is available? What is a lay person to do? Assuming that the question is factual (and not purely valuational or aesthetic), then there is little hope that I can do better than the experts. But I may be able to do at least as well as they.

For what do the experts do when their field does not permit unanimous or majority opinion? They *choose* a particular point of view, aware that the field, as repository of the collective wisdom, is incapable of producing a *decision* in the matter. When there is consensus, it is as if the field as a whole provided the equivalent of an informal decision procedure for the problem at hand. The consensus comes about precisely because the majority of experts apply the same theory to the same problem and obtain the same result. The consensus is obtained, one is tempted to say, *objectively,* in the sense that the field operates like a vast, diffuse decision procedure, which each expert uses to crank out the answer.[5] Thus, the consensus indicates the existence of a paradigm, within the field, for that sort of problem, and that paradigm can be profitably seen as a diffuse decision procedure which explains, in part, the consensus itself. The point is that such *decisions* are not *choices.*

What, then, is a *choice* in this context? It is both the method and the result of fixing one's belief or preference on the basis of some-

[5] Of course, the reader should take this description with a good-sized grain of salt; I certainly do not wish to portray the individual researcher as a simple cog in a decision procedure. One should have in mind, perhaps, the sort of situation where all geometers are agreed on the correctness of a theorem, because each has verified that it is indeed a consequence of the axioms of the system. The collective result is rendered relatively clear by the existence of an algorithm or decision procedure, but it in no way detracts from the individual merit of each geometer. In the messier physical, and even messier social, sciences, algorithms are few and far between. But the consensus, when it does exist, indicates that the members of the community have a very similar outlook on the problem and very similar ways of interpreting and resolving it.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

thing other than a decision procedure, no matter how diffuse. If a decision is based on a more or less objectively agreed-upon method operating on more or less objective data, then a choice is an option that has no such objective method or basis. In an ethical context, choice is often presented as the result of the operation of some special faculty of the human mind or soul. Here I wish to forego any ethical connotations, not because they are unimportant or because they are not germane to the epistemic domain (they are), but because all I wish to bring out is the difference between decision and choice; and, I am quite happy to leave it to others to determine what faculty, if any, is responsible for the function of choice itself.

If the experts do not concur, then I may do as each expert does individually: choose a point of view on the matter at hand. Of course, there is an asymmetry here: whereas the expert may have libraries full of data and a lifetime of experience behind him, I, the layperson, do not, and, even if I did, I probably would not know how to handle it. My choice, though not as considered as the expert's, is still just as legitimate an action as his, for the same basic reasons: though perfect knowledge may be the ideal situation, it is never available, and rarely is the available knowledge even adequate. But the condition of living beings, and of human beings in particular, is such that action must be taken in most situations without perfect knowledge, and sometimes without any knowledge at all. This is precisely what pragmatism is all about. So, if the situation merits action on my part, and the community of experts is divided as to the correct course of action I should take, then I may legitimately act according to my best lights; after all, that is what the expert does when he or she chooses a solution according to his or her best lights.

The point here is most emphatically not that lay opinions are in all cases on a par with those of the expert. For when the community of experts in the relevant field is agreed on a question, that is because their discipline, presumably the best source of knowledge humanity has on the subject, provides the theory and factual basis conducive to decisions as delineated above. It would then, indeed, be irrational for the layperson not to heed (or even not to consider at all) expert opinion. When, however, expert opinion is divided or nonexistent, this means that there is no reliable expert opinion available: expert and layperson alike are left to their own devices. So it becomes rational for the layperson to choose a particular expert as a guiding light for all sorts of reasons ranging from general style to past records of success, to ineffable intuitive leaps, to coin-tossing. What would be irrational in such a situation would be to emulate Buridan's ass, and hesitate before the various options until the sky fell in.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

MICHEL J. BLAIS
Université de Sherbrooke

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

Change in View: Principles of Reasoning. GILBERT HARMAN. Cambridge, Mass.: MIT Press, 1986. viii, 147 p. $19.95.

In *Change of View,* Gilbert Harman endorses a valuable insight of the pragmatist tradition. He insists that one requires good reasons for giving up beliefs belonging to one's evolving doctrine just as one requires good reasons for adding new beliefs to it.[1]

Harman adopts a "principle of conservativism" which declares that one is justified in continuing to accept or believe something fully in the absence of a special reason not to (32, 46). Conservativism is one of two competing tendencies or aims which, he believes, drive the way reasoned revisions of belief ought to proceed. The other is the injunction to maximize coherence.

Although the principle of conservativism is plainly intended to support a symmetry between demands for justifying subtractions and additions to one's body of full beliefs, some of the substantive implications of Harman's use of the epithet seem questionable.[2]

Harman's suggestion that there are several competing aims or tendencies in reasoned revision seems to me right; but it is doubtful that we can usefully characterize them as maximizing coherence and minimizing change. The aims on which we should focus are avoidance of error and maximization of informational value. From the point of view of the inquiring agent, a risk of error is incurred in adding new items to his full beliefs which are not already entailed by them. Hence, the agent needs an inducement to do so. The informational value (including features like explanatory power which may figure in Harman's conception of coherence) provided by adding new items to his full belief should compensate for the risk of error

[1] I have explicitly endorsed this view in my *Enterprise of Knowledge* (Cambridge, Mass.: MIT Press, 1980), p. 2. The idea goes back at least to C. S. Peirce's "Fixation of Belief," *Popular Science Monthly* (November 1877); variously anthologized, e.g., in Peirce, *Chance, Love, and Logic,* Morris Cohen, ed. (New York: Harcourt, Brace & World, 1923).

[2] I would have preferred calling it "the principle of cognitive (or doxastic) inertia." The difference between the Newtonian law of inertia and the Aristotelian law pivoted on whether a force was needed to account for change of place or merely for change of velocity. In the context of discussions of reasoned revisions of belief, we need to identify our presuppositions concerning which kinds of changes in belief may appropriately call for reasoned revision. When someone insists, as both Harman and I do, that one should have reasons for abandoning beliefs one currently adopts, just as one needs good reasons for coming to believe propositions not initially endorsed, there is sufficient analogy with the physical case to warrant the metaphor.

0022-362X/87/8407/0376$00.90 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

incurred. Thus, in adding items to our beliefs, it is the concern to avoid error and not a conservative resistance to change which calls for a justification of the revision.

In removing items from one's full beliefs, no risk of error is incurred; for one takes for granted that there is no error in one's full beliefs—at least not in the sense in which truth and error are judged "relative to the evolving doctrine." And that is the sense of 'error' in which, so I suggest, avoidance of error is a desideratum in inquiry. In giving something up, one will incur a loss of informational value contained in propositions one initially fully believed to be true. Such a loss of informational value is gratuitous unless there is some other inducement to acquiesce in it—such as the prospect that, in subsequent inquiry and belief revisions, one will end up with more informational value than that with which one began. It is an interest in minimizing loss of information valuable as a resource in inquiry, and not a conservative hostility to change, which furnishes the basis of reasoned resistance to abandoning erstwhile settled assumptions.

Nonetheless, Harman sees his topic as being reasoned revision of belief and recognizes that he needs to consider revisions where beliefs are given up as well as cases where they are added. He seems to be turning his back on "pedigree" theories of belief revision, which insist that no belief ought to be retained unless it has been certified legitimate. This is all to the good.

Unfortunately, Harman adopts a view of how normative accounts of reasoned revisions of belief ought to be related to positive accounts of the way agents actually do revise their beliefs which severely undermines this rejection.

Harman contrasts explicit beliefs with implicit beliefs. The former type of belief "involves an explicit mental representation whose content is the content of that belief" (13). Implicit beliefs are related to explicit beliefs in one of the two following ways: (a) as beliefs easily inferable from the agent's explicit beliefs, and (b) as beliefs to which one is committed by one's explicit beliefs. Although Harman does not say so, it is a sensible terminological practice to suppose that the agent is committed to all his explicit beliefs and to all his implicit beliefs of type (a) in addition to the type (b) beliefs. Hence, one can say that implicit beliefs are beliefs to which one is committed in addition to one's explicit beliefs.

Given this distinction, we may consider reasoned belief revision to be a revision either of the set of explicit beliefs or of the set of explicit and implicit beliefs together—i.e., the set of beliefs to which one is committed. The difference is that sometimes an implicit belief may become explicit without any alteration in the agent's commit-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ment. This might be considered a change in the explicit part of the commitment, but it is not a change in commitment. Harman does not consider this matter explicitly. He seems to be concerned with the revision of explicit beliefs while conceding that implicit beliefs may be relevant to the legitimacy of revisions of explicit beliefs. I myself have found it more advantageous to focus on revisions of commitments while ignoring changes in which some erstwhile implicit fragment of a commitment becomes explicit.

Harman wisely refrains from giving an explicit definition of what it is to be committed by one belief to another. Instead, he identifies principles according to which one belief generates commitment to another. (Four kinds of commitment involved in believing *P* in an authoritative way are listed on p. 52.) This is where serious differences emerge between Harman and me.

Recall that type (a) implicit beliefs are not explicitly believed but can be converted to explicit beliefs if it becomes desirable to do so, because they are "easily inferable" from explicit beliefs. It is within the agent's competence to live up to his commitment to such implicit beliefs explicitly when called upon to do so. Type (b) implicit beliefs are not, in general, easily inferable from explicit beliefs because they are not inferable at all—at least so says Harman. They are, instead, presuppositions taken for granted when the agent acquires explicit beliefs. However, these presuppositions may be tacit rather than explicit. Although Harman does not say this in so many words, it appears that these tacit presuppositions can be rendered explicit fairly easily if necessary. Thus, no item can be an implicit belief of either type (a) or type (b) unless the agent is capable of easily rendering it explicit. We cannot include a belief among an agent's doxastic commitments unless the agent has the competence to live up to that commitment explicitly if required to do so. This restriction on commitments is too stringent. To appreciate the point, consider whether a body of beliefs ought to be deductively closed.

Harman denies (and I maintain) that the body of full beliefs to which an agent is committed at a given time ought to be closed under deduction. The argument for Harman's rejection of deductive closure proceeds by first considering it as a condition on explicit beliefs and then exploring its merits as a condition on explicit and implicit beliefs together.

It is obvious that no agent could explicitly believe all the logical consequences of his beliefs as "the logical closure principle" would require if it applied to explicit beliefs. Nor is it desirable for agents to do their best to approximate observance of this principle; for that would entail cluttering up the mind with trivial beliefs, in conflict

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

with a favorite principle of Harman's: "the clutter avoidance principle," which enjoins us against storing trivialities in the mind. (Harman takes the clutter-avoidance principle to be a metaprinciple constraining principles of belief revision; 15.) He defends its use by reminding us of the limited storage capacity of the human agent.

Harman insists that it does no good to apply the logical-closure principle to the union of the explicit and implicit beliefs. He protests, "One cannot be expected even implicitly to believe a logical consequence of one's beliefs if a complex proof would be needed to see the implication" (14).

This complaint is compelling if it is required that implicit beliefs of types (a) and (b) both be easily rendered explicit when the need arises. Thus, Harman's view of commitments presupposes that an agent cannot be said to have commitments to do or to believe or to seek something unless he is capable in advance of living up to the commitment.

There is, however, another view of the matter (see my *op. cit.*, 9–13). Commitment to believing the logical consequences of our beliefs does not entail that we are able to live up to our commitments. It does point to the desirability of developing ways and means of extending our limited capacities to do so. To this end, we promote technologies like paper and pen, the printing press, libraries, and computers because they supplement our memories and increase our computational facility. We encourage the study of logic and mathematics because they help identify the extent of the commitments incurred by the logical-closure rule. And at least one function of psychotherapy is to enhance our capacities to avoid the obstacles to careful thinking engendered by sundry forms of emotional disturbance. The fact that there can be what Harman calls "errors of reflection" (errors of carelessness, inattentiveness, memory, computation) and, indeed, that no one is capable of eliminating them entirely points to an important problem: How can we improve the capacities of agents to live up to the commitments they have endorsed? But it does not necessarily nullify the commitments.

I grant Harman the injunction against cluttering one's mind unnecessarily. But what should be stored "in the head" is a function not only of human interest (Harman is anxious to emphasize this aspect of the matter), but also of the state of the art of information storage and retrieval (a point he overlooks). When written records were hard to come by, specialized reference works were not easily accessible, and neither handbooks of mathematical formulas nor hand calculators were available, scholars stuffed their heads with all sorts of information which, given the current state of technology, no longer

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

seems necessary. The Rabbis used to have assistants who knew the Mishnah by heart and who served the function of published copies of the Mishnaic text. As books and writing of memos became easier, note taking and record keeping substituted for memory. Why retain in the head what one can write down in notes or check in a reference book?

Students of computer technology and artificial intelligence, psychologists and psychotherapists, librarians, book publishers, logicians, and mathematicians all have an important contribution to make to facilitating our capacities to live up to our commitments when called upon to do so. But they can help only if they are furnished with some conception of the conditions required for living up to a commitment.

The logical-closure principle, construed as a constraint on commitments to full belief, does elucidate the character of these commitments in a way which gives just such direction to these helping disciplines. On this view, it is false to suppose, as Harman does, that empirical findings that indicate a persistent tendency on the part of experimental subjects to reason in some specific way furnish good reasons for tailoring our norms to meet these facts. The behavior might reveal, instead, defects in the performance and, perhaps, the capacity of experimental subjects and so point to the need for a technological or therapeutic fix. It is not always easy to decide whether a systematic tendency on the part of human agents to deviate from norms of rationality points to the unacceptability of the norms or to a deficiency in the agents calling for therapy. In the case of beliefs, I suggest as a rule of thumb that when failures are recognizably due to lapses of memory, limitations in computational capacity, inattentiveness, or emotional disturbance, the blame should be placed on the agent and a remedy be sought in therapy. When failures do not seem to be characterizable in these ways, we should at least consider whether the norms are defective.

None of this establishes, of course, my contention that a system of commitments to full beliefs ought to satisfy logical closure. I merely wish to point out that Harman's rejection of a principle of logical closure on the union of explicit and implicit beliefs (i.e., commitments) assumes a conception of what those commitments are which is itself challengeable by those who think that deductive logic has some significant bearing on the question of belief revision. I agree with Harman that principles of deductive logic should not be used to regulate how agents revise their beliefs (construed as their doxastic commitments). Instead, they help define what those commitments are. But, even if principles of deductive logic are not principles of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

belief revision, they remain relevant to belief revision; for they characterize what is being revised.

Harman's view of commitments, as I have indicated, is quite different from this. He seems to hold the view that our commitments ought not to outstretch our capacities to live up to them. With this understanding, he is able to maintain with some plausibility that logic has little relevance to a general theory of the revision of commitments we are capable of living up to. I would grant him his claim, but then suggest that Harman has missed the important tasks to be addressed by those concerned with giving accounts of reasoned revision of full belief. Our full beliefs are intended to define the space of possibilities (I call them "serious possibilities") within the framework of which our judgments of uncertainty, our valuations of prospects, and other appraisals pertinent to deliberation are conducted. Given this understanding of the services that a body of full beliefs is intended to render in deliberation and inquiry and given the idea that, ideally, an agent ought to fully believe that *P* if and only if it is not a serious possibility according to him that *P* is false, in our less than ideal situation, with the less than perfect doxastic instrumentalities we have at our disposal, we may, nonetheless, declare ourselves committed to using one of their perfected modifications. Such a perfected modification, I would argue, ought to be closed under deduction.

The plea just entered for a deductive-closure requirement on commitments to full belief is predicated on a conception of the use to which deliberating and inquiring agents put their views. I do not know what Harman thinks about the functions of beliefs in deliberation and inquiry. Harman does not consider what a network of beliefs is good for and apparently does not recognize the urgency to do so. I suspect this is because his gaze is directed to the role of belief (and desire) in the explanation of action. Still, insofar as the official problem of Harman's book is revision of belief and since the considerations that should regulate such revisions ought arguably to be controlled by the functions of belief states in deliberation and inquiry, some attention ought to have been devoted to identifying the functions that belief states are to perform in deliberation and inquiry.

I began by applauding Harman's insistence that we need to have good reasons for abandoning full beliefs just as we need to have good reasons for adding to our full beliefs. Harman appreciates that such a view stands in opposition to assorted "foundationalist" views which insist that one should jettison every full belief for which one lacks a good reason. Indeed, it stands in opposition to the more general

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

category of "pedigree" theories of belief revision, knowledge, etc., which include forms of naturalism that require the good reasons to be legitimating causes. Unfortunately, Harman's understanding of the character of the principles of commitment which serve to define the set of explicit and implicit beliefs subject to revision is a vital ingredient in his virtual surrender to advocates of pedigree theories.

Harman raises the white flag when he strengthens his principle of conservativism by adding a "principle of positive undermining" (39). According to this principle, one should stop believing *P* whenever one positively believes that one's reasons for believing *P* are no good. Harman takes this principle to characterize a commitment that the agent has in virtue of explicitly believing *P*. The agent is committed to believing that he has or had sufficient reasons for believing *P* which did not rely on any false assumptions (43/4, 52). Observe that this proposition is not "easily inferable" from *P*, so it is not an implicit belief of type (a). The agent is, nonetheless, committed to it as a tacit presupposition of his believing that *P*. In this sense, the principle of positive undermining offers a partial characterization of the agent's doxastic commitments in Harman's scheme, just as the principle of logical closure does in mine. And, just as Harman denies that our doxastic commitments ought to be regulated by deductive closure, I deny that they should be regulated by the principle of positive undermining.

Harman contrasts the use of this principle, which is congenial with his "coherence" theory of belief revision, with the "foundationalist" approach. Harman's foundationalist endorses a "principle of negative undermining" according to which one should stop believing *P* whenever one does not associate one's belief in *P* with an adequate justification (39).

Presumably the principle of negative undermining urges the jettisoning of beliefs that the principle of positive undermining would not recommend abandoning. To explain this, Harman appeals to examples in which an agent is informed of unreliability or error in the data on which some belief *P* was initially acquired. Even if the agent is informed of the error, the agent may have failed to keep track of the justification of his belief—a practice which, according to the clutter-avoidance principle, might be eminently sensible. Nonetheless, the agent might continue to believe explicitly that the belief has reasons and some of them are "good." In that event, the principle of positive undermining fails to legislate removal of *P* from the agent's beliefs. But the principle of negative undermining requires that the agent identify a specific adequate justification for *P*, rather than merely that there be one. Hence, *P* should be removed.

It is unfair and misleading to saddle the foundationalist with the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

view that agents ought to keep a record in their heads of all the justifications of their beliefs. They need not explicitly believe all the information such a record would contain. They can, nonetheless, be committed to this content. Of course, Harman would insist that one should not be committed to such records unless the agent is able to recover them easily; and in the case of memory he might insist that this entails retaining them "in the head."

But, even if we could not be committed to an item of information unless it is easily recoverable by us, it is far from obvious that the item needs to be stored in the head or somehow derivable from such a store. As already pointed out, we have all kinds of helpers to facilitate recovery.

I would go further. Even if this extended capacity is limited, we regard as desirable the improvement of the technology called for to enlarge that capacity. And this attitude sustains the idea that our commitments outstretch our capacities.

Thus, foundationalist views are not to be dismissed simply because of the finitude of our memories and the desirability of reserving the use of memory for important matters. Foundationalists who are not trapped in Harman's narrow conception of commitments can finesse Harman's objections by demanding that the agent be convinced that all the full beliefs he adopts at one time have good reasons whether the agent is able to easily identify them or not. But such a position is virtually equivalent to endorsing the principle of positive undermining as Harman understands it.

Thus, Harman's ability to deploy the differences between the principles of positive and negative undermining depends crucially on the same vision of doxastic commitments which grounds his rejection of deductive closure as a condition on doxastic commitments. And this vision is itself predicated on a conception of the relation between doxastic commitment and doxastic competence (capacity to live up to doxastic commitment) which ties the former very tightly to the latter. As a consequence, adequate accounts of reasoned belief revision need to wait for guidance from psychological, biological, and sociological investigations into human doxastic capacities.

Harman uses the conception of the relation between competence and commitment to argue that studies of reasoned belief revision ought to focus on revision of full belief and not on degrees of belief probabilistically construed. (Harman ignores nonprobabilistic conceptions of degrees of belief altogether.) And, in his discussion of reasoned revisions of intentions, he deploys essentially the same approach covering many topics including, as one might expect, a dismissal of expected-utility maximization.

Harman is right to insist that normative questions concerning how

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

we ought to revise our beliefs and plans should not be detached from positive investigations into our actual conduct.

Nonetheless, psychology and computer science do best when they furnish us with therapeutic and technical assistance so that we can do a better job of living up to our commitments. We should not always trim our commitments to accommodate the latest empirical findings.

ISAAC LEVI

Columbia University

The Significance of Philosophical Scepticism. BARRY STROUD. New York: Oxford University Press, 1984. xiv, 277 p. Cloth $27.50, paper $10.95.

In this very thoughtful book, Barry Stroud has not sought to defend or to refute the claim of philosophical skepticism "that we can know nothing about the physical world around us" (viii). Instead, his aim has been to explain the "significance" of this claim. Understanding just what this claim could mean and discerning what its falsity would entail will place us in a better position, he believes, to grasp "something deep or important about human knowledge or human nature or the urge to understand them philosophically" (ix). Of course, many epistemologists believe that they have put the skeptical problem behind them. But Stroud rightly insists that, even if that is where this problem belongs, the reasons why this should be so remain unclear.

The book falls into seven chapters. Beginning with Descartes's statement of the skeptical problem in the *First Meditation,* Stroud turns in the next chapter to the relation between skepticism and "everyday" knowledge claims and to J. L. Austin's verdict that skepticism rests on a misuse of the word 'know'. There follow four chapters discussing the antiskeptical positions of G. E. Moore, Immanuel Kant, Rudolf Carnap, and W. V. Quine. The last chapter deals with the diagnoses of skepticism offered by Stanley Cavell and Thompson Clarke. Plainly this is an ambitious book. At the same time it is also a modest one. Stroud acknowledges that he has not arrived at a precise formulation of the moral that philosophical skepticism should teach us (ix). The strength of his book lies instead in the acuity with which he brings out what is relevant in the different responses to skepticism, what are the obstacles they face, and what are the fundamental assumptions about truth and knowledge which underlie

0022-362X/87/8407/0384$00.90 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

this debate. I shall present some major criticisms about how Stroud approaches the problem of skepticism, but they should not obscure how successful this book is in showing that skepticism remains a serious and fruitful philosophical problem.

The chapter on the *First Meditation* examines the way in which the skeptical problem arises. The general doubt whether I can know from the senses anything of the world apart from my "ideas" about it results, Stroud argues (7f., 14), not from the mere fact that sometimes I make perceptual mistakes, but rather from the threefold conviction that (a) if a certain possibility (e.g., I am dreaming) were true, none of my perceptual reports could then count as knowledge, (b) only if I knew that possibility to be false could I know that my perceptual reports are true, and (c) I can never know that that possibility is false. Stroud is right to say that, since (a) is indisputable and since the plausibility of (c) depends on (b) (any test to tell whether I am dreaming could be decisive only on the assumption that I was not dreaming about carrying it out), the key to the skeptical argument is (b) (23). The formulation that Stroud gives to this principle, which I shall designate as "K," is: if someone knows that *p*, he must know the falsity of all those things which are incompatible with his knowing that *p* (29). Stroud holds that, if K is indeed a requirement on knowledge, then skepticism will be ineluctable (30).

Chapter Two examines Austin's claim that skeptics distort the meaning of the word 'know'. A possibility (e.g., my dreaming) whose occurrence would defeat a knowledge claim is relevant, Austin claimed (52), only if there is some special reason for thinking that it indeed obtains. According to Stroud (48), Austin thus rejects K. Austin based his objection on how we ordinarily use the word 'know', for ordinarily I would treat the possibility that I am dreaming as irrelevant or outrageous unless I had some special reason (e.g., it was the middle of the night and I felt groggy) to think I might be dreaming. Against this objection Stroud first observes that, generally, the ordinary conditions for reasonably asserting "I know that *p*" need not coincide with its truth conditions or "meaning" (57ff., 75 fn) any more than they must in the case of any other proposition. Thus he finds Austin's charge of meaninglessness unfounded. Next he claims that K will form part of the meaning of "I know that *p*" (63, 76, 78–81) if we adopt an "external" view of the relation between knowledge and the world (81)—if, that is, we hold that the world we want to understand "is the way it is quite independently of its being known or believed by us to be that way" (77). For the idea that knowing that *p* requires knowing the falsity of only a proper subset of the things incompatible with knowing that *p* would imply that the world of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

which *p* is true is relative to the framework of that subset. Stroud suggests that we do not ordinarily take K as a requirement on knowledge, only because of "practical" considerations (time, effort, ingenuity, need for action) distinct from the "theoretical" aim of knowing the truth about this world (62f., 65f., 71). His conclusion is that the external standpoint must thus lead to skepticism (82, 119).[1] I shall come back to this argument at the end of the review.

In the next three chapters Stroud investigates the different efforts of Moore, Kant, and Carnap to overcome skepticism by refusing to adopt, in effect, this external standpoint. Focusing on Moore's famous "proof of an external world," and on the three necessary conditions of a successful proof which Moore laid down—its premises ("Here is a human hand" and "Here is another human hand") are different from its conclusion ("Two human hands exist"), the premises are known by Moore to be true, and the conclusion follows from the premises—he claims that the proof does not fail simply because, as we might naturally think, the second condition is not satisfied. Rather, it fails, Stroud claims, because it tries to establish a thesis, "there exist things apart from our minds," by appealing to particular items of knowledge ("here is a human hand") the grounds for which have themselves been challenged by the general considerations (such as principle K) that the skeptic raises (108–113). By simply ignoring any doubt addressed to the totality of our knowledge claims, by responding to the skeptic's question from a position within a body of given knowledge, Moore thus held on, says Stroud, to an "internal" standpoint (117–119). Since his proof fails to refute the skeptic, Stroud concludes that either "Moore's assertions of knowledge are not true or . . . they do not even contradict philosophical scepticism" (127).[2] The second horn of this dilemma is the position that Stroud finds elaborated, in different ways, by Kant and Carnap (cf. 157, 159, 176).

In a very important respect Stroud's account of Moore seems to me inadequate. A distinctive feature of Moore's epistemology was his conviction that (first-order) knowledge claims like those in these premises are more certain than any general (second-order) principle about the nature of knowledge such as those which fuel the doubts of the philosophical skeptic. Given this conviction, he could dismiss the skeptical doubt as resting on reasons less certain than his ordinary

[1] This external view expresses, of course, what Bernard Williams has described as "the absolute conception of reality." See his *Descartes: The Project of Pure Enquiry* (Harmondsworth: Penguin, 1978), ch. II.

[2] Inexplicably and regrettably, Stroud says next to nothing (93) about Wittgenstein's critique of Moore's antiskepticism in *On Certainty*.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

knowledge claims. Stroud mentions this Moorean argument (106). But he objects that, so far, it has not been shown that a general assessment of our knowledge claims does not lead to the skeptic's conclusion and, thus, that Moore should not have leaned upon his knowledge claims to dismiss that conclusion (111). This reply, however, simply *ignores* Moore's insistence that his (first-order) knowledge claims are more certain than any (second-order) principles involved in a general assessment of such claims. As a result, Stroud's objection is misconceived, if it is advanced in the spirit of the skeptic. "Skepticism," of course, has meant many things, not all of them very interesting. But the most important skeptics, from Sextus to Hume, have recognized that, if the skeptic's challenge is to be relevant and compelling, his business cannot be to substitute his own notion of knowledge for the notion invoked by those who claim to know things. (After all, the philosophical skeptic cannot consistently claim that he knows what knowledge of real objects is.) Instead, he must uncover the defects in their notion of knowledge, showing that it is self-contradictory or conflicts with what they claim to know. One reason, therefore, why Stroud's critique of Moore misfires is that he overlooks the skeptic's necessarily parasitic strategy of internal demolition. Instead of showing some conflict among the things that Moore himself believes, he dismisses Moore's position on the basis of his own views about knowledge. As I shall show, this is not the only case where this sort of mistake occurs. Another, related reason why Stroud is unfair to Moore is that his repeated appeals to "our ordinary conception of knowledge" (31)—as though this were something timeless and homogeneous—blind him to how, in fact, ideas of knowledge vary and, therefore, to what is distinctive about Moore's. This view, too, leads to distortions elsewhere in the book as well. For example, the skeptical arguments of the *First Meditation* were aimed not at "our ordinary notion of knowledge" as Stroud maintains (31, 43, 55), but rather at a Scholastic notion of knowledge which, in the subsequent *Meditations,* Descartes sought precisely to replace with a new notion that he believed immune to such arguments.

Stroud's chapter on Kant is one of the best expositions in this book. Kant maintained that, if the epistemic priority of ideas over real objects is admitted, then skepticism is unavoidable (131/2, 141, 144). So his task was to show that this Cartesian thesis of epistemic priority is wrong and that we have direct knowledge of real objects which is not inferred from a prior acquaintance with ideas. Moreover, he thought that the only way we can know that such noninferential knowledge of real objects is a necessary feature of our under-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

standing is to know it a priori, and that we can know it a priori only if having this sort of knowledge of real objects is a condition we place upon our experience (148/9, 154/5). Thus, as Stroud observes, Kant's refutation of skepticism requires the doctrine of *transcendental idealism,* the view that in a certain respect the real objects of which we have knowledge are dependent on us. This doctrine is a response, Stroud notes, to the "external" question about the relation between our experience and a world independent of it (150, 156–159), a response which denies that our knowledge can be of such a world. *Within our experience,* or "internally," we can distinguish between our ideas and objects and recognize that the latter are not dependent (causally) on the former; this is Kant's doctrine of "empirical realism" (150/1). So Kant's antiskepticism consists in the supposed compatibility of these two doctrines: he asserts the possibility of knowledge within experience, while conceding that from an external viewpoint skepticism would be correct.

Against this position Stroud makes several objections, two of which are very important. He argues that Kant's position is unsatisfactory because, however much it differs from "empirical idealism" (the skepticism of the *First Meditation*), it leaves us "with knowledge that is too centered on us, too subjective, and in that respect not what we originally aspired to" (167). I find this objection unfair for the same reason I found Stroud's critique of Moore unfair: he rejects Kant's notion of knowledge simply because it differs from what he supposes is our "ordinary" or "original" notion. More convincing is another of Stroud's objections, which is also familiar. Transcendental idealism consists in the view that we place conditions on our experience which make objects dependent on us. But, Stroud points out, Kant was committed to two other views that conflict with this: intelligible thought is possible only within the limits of possible experience, and within these limits we recognize objects as being independent of us (161). This is the right sort of skeptical objection to make to Kant's theory of knowledge, because it uncovers an internal inconsistency.

In the next chapter, Stroud shows how Carnap's "Empiricism, Semantics, and Ontology" tried to circumvent this difficulty by means of his verificationist theory of meaning. Any general explanation (such as transcendental idealism) of how we can legitimately claim to have knowledge of objects within the limits of possible experience is now meaningless precisely because it itself goes beyond those limits (168/9). But, similarly, skepticism is impossible, since, for any proposition about objects to be meaningful, it must be, in principle, empirically verifiable or confirmable (173). Carnap agreed

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

with the skeptic that, if our knowledge must be of a world as it is in itself, independent of all possible experience of it, then knowledge is impossible. But he was no skeptic, Stroud observes, because this "external" viewpoint and so also the skeptic's claim that knowledge is impossible from this viewpoint are strictly meaningless (177–179). Only internally, within the limits of experience, or—as Carnap said—within the framework of the thing language, can we meaningfully talk about objects. The question whether we should use the thing language, instead of just the sense-datum language, is not—as both what Stroud calls the "traditional philosopher" (i.e., one committed to the external viewpoint) and the skeptic suppose—a "theoretical" question, to which the answer could be "there really are things or objects." Rather, for Carnap, it is a "practical" question of deciding which framework is more useful (186).

In Stroud's view, Carnap did not really escape Kant's difficulties. Since, for Carnap, no proposition about mountains in Africa is true or false apart from our adopting the "thing language," Stroud dismisses this as an "unacceptable idealism" (193). This objection is no better, I suggest, than his similar rejection of Kant's position as "too subjective." But his more convincing criticism of Kant also finds a parallel. Carnap's distinction between "theoretical" and "practical" questions—which, like transcendental idealism, constitutes a general explanation of how knowledge is possible "internally"—violates the verificationist principle of meaning. For the thing language and the sense-datum language cannot be understood as practical alternatives except in terms of some common task (e.g., coping with experience) they are to perform, and that common task must thus be understood, not internally to either of these frameworks, but apparently "externally" (191, 196 ff). Indeed, as Stroud should have mentioned, the verification principle of meaning itself, being neither analytic nor empirical, oversteps the limits of intelligibility that it itself lays down. Carnap's position contradicts itself.

Rejecting Carnap's distinction between the theoretical and the practical, Quine insists that the thing framework, no less than any other theory, is accepted in the light of what else is already known. Stroud devotes his next chapter to how Quine, in thus avoiding Carnap's inconsistencies, has dealt with the skeptic's doubt.[3] Dispensing with Carnap's dualism has allowed Quine to claim that philosophy, the systematic reflection upon science (or empirical inquiry in general), is continuous with science, instead of being based outside

[3] The key texts are W. V. Quine, *The Roots of Reference* (LaSalle, Ill.: Open Court, 1974), pp. 1–4; and "The Nature of Natural Knowledge," pp. 67–81, in S. Guttenplan, ed., *Mind and Language* (New York: Oxford, 1975).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

it (in transcendental argument or conceptual analysis). This has led him to hold that the skeptical challenge arises from science itself: science informs us of the gap between our meager sensory input and what it tells us that we know about the world, and the skeptic, in harping on this gap, tries to use science to refute science (224–228). But, Quine replies, we can counter this reductio argument by noting how science tells us this gap is crossed (224, 231/2).

Stroud presents two objections to Quine's approach. One is that, contrary to Quine, we cannot regard all our beliefs about real objects as posits or hypotheses constructed from sensory inputs at the same time as we use such beliefs to explain how we can come to know about real objects on the basis of sensory inputs (242/3). I must say that I fail to see the difficulty. Surely it is not impossible for a system of beliefs to explain its own possibility. The reason for Stroud's objection seems to be the odd idea that we cannot regard a belief about objects as at once true and inferred, or hypothesized on the basis of sensory inputs (247). Stroud's other objection is that we are not free to use scientific beliefs to counter the skeptic's challenge, because it is precisely those beliefs which the skeptic has put into doubt (229). This riposte seems to me to show the same misunderstanding of skepticism which I have indicated before. The skeptic's strategy cannot be to impose standards from without; his aim must be to show that someone who claims to have knowledge of the external world is caught in internal inconsistency. This is what Quine himself sees with perfect clarity, and the import of his appeal to scientific beliefs to counter the skeptic is that science can be shown not to harbor any such inconsistency. (Whether Quine succeeds is another question. We need to learn, for example, a lot more about his criterion of knowledge, and it is also unclear whether his position coheres with his doctrine of the indeterminacy of translation, which Stroud does not discuss.)

Stroud's last chapter, dealing with Cavell and Clarke, adds little to his previous discussion. But he does reiterate his earlier claim that the "external" standpoint, if it is intelligible, must lead to skepticism (273/4). In this connection lies the "significance" of philosophical skepticism, according to Stroud. To this matter I now want to turn.

In Chapter Two, Stroud argued that the principle I have labeled K belongs to the meaning of "I know that *p*," if we think of the world "externally," as being the way it is quite independently of its being known or believed by us to be that way. This is so; but I do not see why K itself must lead to skepticism. The reason is that K says a lot less than Stroud apparently believes it does. If we are to be entitled to claim that we know that *p*, K stipulates only that we must also be entitled to claim that we know the falsity of all those eventualities

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

incompatible with our knowing that *p*. So it leaves open the conditions under which we may claim to know the falsity of those other eventualities. For example, these conditions might be that, in the absence of a positive reason to suspect that any of those eventualities might be true, we may take for granted that it is false. In this case, K would be compatible with Austin's notion of justification after all. And, thus, thinking of the world "externally" would not lead to skepticism.

I believe that Stroud thinks otherwise because he imports into K more than it actually claims. For him, K means that I may claim to know that *p* only if I can give a reason for the falsity of each of those eventualities incompatible with my knowing that *p*. Thus, he simply reads into K the idea that, instead of needing a reason to take seriously the possibility of dreaming (as Austin would say), we need a reason to show that this possibility is false.

How should we choose between these two versions of K? At times Stroud (like Austin) seems to think that any principle operative in our "ordinary conception of knowledge" must be maintained, for "we have no notion of knowledge other than what is embodied in those practices and procedures" (31). But I doubt that there is any such "ordinary conception" timeless but determinate enough to imply just one of these versions. As I suggested earlier, Stroud tends to underestimate how much conceptions of knowledge vary, both historically and within our present everyday and scientific practices. Instead of resorting to conceptual analysis, we would do better to look at these different versions of K as alternative conceptions of how the goals of inquiry are to be ranked. Both are committed to the idea that our knowledge claims purport to be of an independent world. But Stroud's version implies that it is always better to avoid a possible falsehood about this world than to accept something as true about it. (These two goals are distinct; for, roughly, if we were interested only in avoiding falsehoods, we should claim to know nothing, and if we were interested only in acquiring truths, we should claim to know everything.) The other version does not place the goal of avoiding falsehoods always above the goal of acquiring truths. More open to risk, it makes the first goal pre-eminent only when there is a positive reason for suspecting that an eventuality might be true which is incompatible with our knowing that *p*. Now the fact that one of these views of the goals of inquiry condemns us to the impossibility of knowledge seems to be sufficient reason to adopt the other, if it does not lead to skepticism. (This second version of K certainly calls for more refinement, and there are many more specific forms it might take, involving different ways of trading off those two goals.)

Stroud says too little in this book about the goals of inquiry. Al-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

though he does discuss the connection between skepticism and the aim of acquiring knowledge of an independent world, he fails to distinguish between the two component, but distinct goals of acquiring truths and avoiding falsehoods about this world, and so he does not consider the question of their ranking and of tradeoffs between them. A closer concern with the goals of inquiry would have led him, I believe, to avoid the two most important errors I have imputed to him. It would then have been more obvious that the skeptical challenge is the parasitic claim that inquiry, as a goal-directed activity, cannot achieve the goals it sets itself. And it would also have been clearer that the forms of such inquiry, or conceptions of knowledge, are as various as the mutual adjustment of means and ends permits.

CHARLES LARMORE

Columbia University

Portraying Analogy. J. F. ROSS. New York: Cambridge University Press, 1981. xi, 244 p. $29.50.*

During its medieval heyday, the theory of analogical predication provided an attractive account of the meaning of divine predicates which avoided the two extremes of complete equivocation and univocality. The analogical shifts of meaning which were claimed to occur in divine predications were claimed, furthermore, to be part of a general linguistic phenomenon of analogy (e.g., 'drop a book/friend/hint'), and Thomas Aquinas and Tommaso Cajetan sketched systematic accounts of this larger phenomenon. Since the end of the fifteenth century, however, the field of analogy has lain largely fallow. There have been various historical studies, oriented toward philosophy of religion, but little has been done to explore the consequences of analogy for the general theory of meaning. In this detailed, dense, and sometimes difficult book, James Ross has now attempted to bring analogy back to the attention of philosophers, both because he believes that the analogy phenomena have "profound consequences" (x) for contemporary philosophy and because the classical theories rest, in his view, on unacceptable conceptions of meaning and analogy.

The specific question about meaning raised by the "analogy phenomena" is "what determines which meaning a word [token] has in a

* I wish to thank Akeel Bilgrami, Jonathan Malino, and Charles Parsons for helpful comments on an earlier draft of this review.

0022-362X/87/8407/0392$00.60 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

given context, and what accounts for its differences from one occurrence to another" (6). Ross's answer rests on two main theses. First, he holds that linguistic meaning is something "inherent" or "intrinsic to the language" and not, as the classical theories held, something "extrinsic" to it, e.g., a mental entity or the thing referred to (11). Second, he claims that the "analogy phenomena" must be regarded as a species of a generic synchronic linguistic regularity that he calls the *differentiation of meaning* (20)—something the classical theories also failed to see. The first six chapters of the book are devoted to these criticisms of the classical theories; general explications of meaning differentiation and Ross's "model" of meaning; a detailed typology of five basic "meaning-relations" that fall under the genus of differentiation, ranging from homonymy ("mere equivocation") through various modes of polysemy, including two kinds of analogy, "simple" metaphor, and paronymity; and an account of "figurative discourse." In the last two chapters Ross draws some consequences from analogy for the cognitive significance of religious discourse and for the philosophical practice of truth-conditional analysis.

Ross's first main thesis, which harks back to the structuralist theory of semantic fields, is that the linguistic meaning of a word (token) *W* is the set of "affinities and oppositions" (12) that *W* bears to other words in the language.[1] This set of relations can be *modeled* by *W*'s "predicate scheme": the set of "meaning-relevant" words in *W*'s "paradigmatic set"—those words which, when substituted for *W*, preserve the "acceptability" of its containing sentence in its discourse environment (54–70). Examples of such words are *W*'s "near" synonyms, contraries, and determinates/ables. Thus differences of predicate scheme "constitute" all differences of *predicative* meaning (66)—though not quite *all* differences of meaning; e.g., they leave out differences of predicate mode (67, 128).

Now, this "model" may provide Ross with an apparatus for describing and categorizing many subtle differences of meaning, but it furnishes no explanation of those differences. Consider the "near synonyms" in *W*'s predicate scheme. Philosophers used to try to explain when two terms were synonymous in terms of sameness of meaning; here Ross tries to explain when two tokens have or do not have the same meaning in terms of their substitutable near synonyms. But, obviously, this will provide us with an *explanation* only if we are given an independent criterion for synonymy which does

[1] For a helpful introduction to this branch of linguistics, see John Lyons, *Semantics*, 2 vols. (New York: Cambridge, 1977), vol. 1, pp. 230–235.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

not circle back to the explanandum. However, when Ross turns to explicating near synonymy, he explicitly appeals to substitutions that yield "replacement sentences *approximately the same in meaning* as the original" (83f., my emphasis)—bringing the circle to a close. Even more puzzling, he next says that "there is nothing to be gained from trying to amplify" (84) this explication, leaving the reader inevitably wondering what his account was all along meant to show.

Even as a descriptive apparatus, moreover, Ross's model faces a serious problem: its uncritical use of intuitive judgments and, especially, of the "theoretically primitive" (58) notion of "acceptability" which is basic to his definition of a predicate scheme and, hence, to his claim to be giving an account of *linguistic* meaning. As Ross himself acknowledges (e.g., 10f., 55, 66, 79–84), many factors can and do affect our intuitive judgment that a sentence uttered on a given occasion is (un)acceptable. Some of these factors are genuinely linguistic or grammatical, but many are not, e.g., memory or processing limitations, matters of conversational appropriateness, or context-dependent constraints such as shared presuppositions. But, without actually separating out the truly linguistic from the others, Ross cannot defend his claim to be explaining a purely *linguistic* sense of meaning. Although he attacks intentional, mentalistic, and ontological accounts of linguistic meaning, he fails to take to heart the distinctions between language and its use and between an interpretation that one knows solely in virtue of one's linguistic competence and one that depends also on one's extralinguistic beliefs or other cognitive faculties. Again, this is all the more puzzling because Ross elsewhere (e.g., 182f.) indicates that he is well aware of these distinctions; yet he somehow fails to see their direct relevance to his claims.

With this model of meaning, Ross proceeds to develop his account of differentiation. Beginning with the linguistic "universals" (9): that all tokens of a word have the same meaning unless they differentiate and that they "resist combining unacceptably in their linguistic environment" (10), Ross says that, when a token does occur in a potentially unacceptable combination, it undergoes "step-wise adaptations" in its meaning to "fit" (6) itself to its context, resulting in different meanings in contrasting contexts. There are always, though, alternative ways in which the constituents of such a sentence can differentiate to render the whole acceptable. Therefore, Ross introduces "dominance-relations" among the constituents to explain why some but not other words differentiate, and which adaptations they make. Disappointingly, however, aside from some examples and hints about the causal properties of tokens (39, 49), Ross gives al-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

most no explanation of how these relations work. He is certainly right that these cases exemplify *some* linguistic regularity that is not simply "up" to the speaker. But it is doubtful that he has explained anything by saying that "whether or not [the speaker] knows [the laws of differentiation], his words obey them: as objects fall, so words go together to make what sense they can" (53).

Furthermore, although the general contour of this idea of differentiation is familiar enough, it is not obvious that all five of Ross's "meaning-relations" really do belong to one genus—a claim which he emphasizes and which he takes to be a signal contribution of his account. In fact, Ross equivocates between two characterizations of differentiation which suggest two distinct kinds of phenomena. On the first, weaker characterization, differentiation is *disambiguation:* with each word (type) there is correlated a number of meanings—multiple predicate schemes to which the word's tokens are "indifferent" out of context—(at least) one of which is then "*selected* in context" (78) for each token "by the *exclusion* of others because the resulting sentences would be unacceptable in the environment" (50, 128). On the second characterization, tokens that differentiate "generate new meanings" (14) and "make room for new ideas" (49); i.e., they *create* new—"endless new" (49)—meanings that they do not have outside the contexts in question. These two characterizations point to two different types of "meaning-relations," the first exemplified by homonymy, the second by metaphor. As many recent philosophers have argued—and as Ross should agree, given his own emphasis that differentiation is a synchronic phenomenon—it is only an accident that the two meanings of a homonym, say, 'bank', happen to belong to one sound/inscription.[2] After all, besides their spelling and sound, what is there *linguistically* or *semantically common* to tokens of such homonyms *from which* their different meanings are claimed to differentiate? When a hearer encounters one of these tokens, his task is simply disambiguation: deciding which of a number of alternative types (individuated by their meanings) to pair with that token on that occasion. On the other hand, the differ-

[2] Ross's identity conditions for "words"—"Roughly, they are the same if spelled and sounded the same" (34)—have even more counterintuitive consequences. Because they leave out any relativization to a language, it would seem that two tokens —one occurring in a language with one meaning, the other in a completely unrelated language with an unrelated meaning—*if* pronounced and spelled the same (and substitutable for each other *salva* meaningfulness), should nonetheless count as the same word. In fact, it will be very difficult to produce examples of this sort because almost all actual languages are phonetically distinct; but, to whet your intuitions, consider 'ashtray' in American English and pigLatin: one word or two? (Thanks to Jerry Sadock for the example.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ent "meanings" that are metaphorically expressed, at least on some occasions, by tokens of one word fit the bill of meaning generation. Here the hearer's task is a matter not of selecting from a predetermined set of alternative meanings but of computing a previously unknown semantic value, given what is often a contextually specific and novel set of variables. Most examples of polysemy fall somewhere between these two poles.

Let me now turn to some consequences Ross draws from his theory of analogy. In chapter 7 he uses his distinction between equivocal (differing in meaning) and merely equivocal (unrelated in meaning) tokens to rebut various versions of the charge that religious discourse has no cognitive content. Instead, he proposes, religious discourse (like legal discourse, of which he has a fascinating discussion in chapter 8) is "craftbound" language; i.e., its use requires the speaker to learn additional conditions beyond those governing its "unbound" (ordinary) use with which it is equivocal, conditions which "cannot be learned without participatory experience in the craft" (166). This is a promising idea, but the real implications Ross sees in analogy run much deeper. In chapter 8 he discusses "constraints" (179) that analogy imposes upon the process of analysis, and, from the start of the book, he asserts that it "refutes the assumptions that sentence meanings are the syncategorematic 'sum' of independent word/morpheme meanings [the "compositionality principle"] . . . and that sentences have truth conditions that are their senses" (1). *If* Ross is right about these bold claims, the analogy phenomena surely do deserve much more attention than philosophers have heretofore given them.

Ross's actual arguments do not, however, support these conclusions. For one thing, what he shows is often rather different from what he claims to show. Thus he never shows that truth conditions are not "senses" of sentences; he shows, rather, that it is *utterances* or *statements* that have truth conditions, which cannot be identified with the linguistic meanings of the *sentences* uttered—incidentally, for now-familiar reasons that are independent of the rest of his theory. Furthermore, even granting that not sentences (types) but utterances (tokens) have truth conditions, *that* qualification hardly undermines, as Ross claims, "the key assumption of the 'Fregean' position" (185).

Similarly, the "constraints" that analogy is claimed to place on truth-conditional analysis add up to the condition that we should not take it for granted that a predicate, e.g., 'know', will always have one univocal meaning in every analyzed context, e.g., regardless of what we substitute for *S* or *X* in the schema '*S* knows *X*'. True; but this is

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

simply the instruction to beware of fallacies of equivocation in one's *use* or *practice* of truth-conditional analysis. It is no objection to truth-conditional analysis itself.

As for "refuting" the compositionality principle, nowhere in the book can I find an *argument* to this effect; perhaps Ross believes the reader will be persuaded simply by examples of words with multiple differentiated meanings, e.g., his eighteen meanings of 'charge' (100). In any event, his version of the principle (quoted above) is close to a straw man. The precise formulation of compositionality is, in fact, the subject of much current debate, but no one holds that the meaning of the whole is simply the *syncategorematic sum* of the meanings of its parts. At the very least, an adequate formulation of the principle must acknowledge the important role in determining the meaning of the whole played by the way in which the parts are syntactically combined.[3] Here, as in his general treatment of sentences as nothing more than "combinations" of words, Ross completely ignores syntax. This is especially unfortunate because even minimal attention to syntactic conditions would have lightened some of the heavy burden of Ross's complicated semantics.

This book, in sum, is not easy reading. Not only the trees, but the branches and leaves often obscure the forest. Ross has a way of using technical terms pages before he explicitly introduces them, and key notions are discussed in several places with no attempt to draw them together. But the most unfortunate fault is Ross's idiosyncratically chosen terminology and theoretical framework. For these set him so far apart from almost all other contemporary philosophers of language that others will inevitably be prevented from learning from, and building on, his insights and his often subtle linguistic intuitions. Although Ross's aim to bring analogy back into the mainstream is admirable, his means undermine his own good intentions.

JOSEF STERN

The University of Chicago

[3] See Barbara H. Partee, "Compositionality," in Fred Landman and Frank Veltman, eds., *Varieties of Formal Semantics* (Dordrecht: Foris, 1984), pp. 281–311.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ANNOUNCEMENTS

Leigh S. Cauman—logician, teacher, editor, and conscience—is retiring as Managing Editor of the Journal of Philosophy after twenty-six years in that position. During this long period, her ability to allow freedom of expression to her writers, while at the same time demanding the highest literary probity and grammatical exactitude, earned her an unsurpassed reputation for editorial excellence. This publication would not have become what it is without her direction and grace, and the profession is in her debt. Although her colleagues will miss her in the day-to-day affairs of the Journal, they look forward to her continuing guidance on the Board of Trustees.

The trustees and editors of *The Journal of Philosophy* announce, with some reluctance, an increase in subscription prices. As of January 1, 1988, these will be as follows:
$40/year for institutions
$25/year for individuals
$20/year for students and retired/unemployed philosophers
$6/year for postage outside U.S.A.

NEW BOOKS: ANTHOLOGIES

O'HARA, DANIEL T., ed.: *Why Nietzsche Now?* Bloomington: Indiana UP, 1985. xii, 441 p. $27.50.

O'MEARA, DOMINIC J., ed.: *Platonic Investigations*. Washington, D.C.: Catholic University of America Press, 1985. x, 271 p. $31.95.

OVERING, JOANNA, ed.: *Reason and Morality*. London: Tavistock Publications, 1985. 277 p. Cloth $35, paper $15.95.

PANICHAS, GEORGE E., ed.: *Marx Analysed: Philosophical Essays on the Thought of Karl Marx*. Lanham, Md.: University Press of America, 1985. xxiii, 308 p. Cloth $26.50, paper $14.25.

PAUL, ELLEN FRANKEL, *et al.*, eds.: *Nuclear Rights/Nuclear Wrongs*. New York: Basil Blackwell, 1986. xiii, 222 p. $24.95.

PENNOCK, J. ROLAND, and JOHN W. CHAPMAN, eds.: *Justification: Nomos* XXVIII. New York: NYU Press, 1986. xv, 368 p. $42.50.

PITT, JOSEPH C., ed.: *Change and Progress in Modern Science*. Boston: D. Reidel, 1985. vii, 214 p. $37.50.

POTTER, NELSON T., and MARK TIMMONS, eds.: *Morality and Universality: Essays on Ethical Universalizability*. Boston: D. Reidel, 1985. xxxii, 312 p. $48.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

PYLYSHYN, ZENON W., and WILLIAM DEMOPOULOS, eds.: *Meaning and Cognitive Structure*. Norwood, N.J.: Ablex, 1986. xi, 264 p. $39.50.

RAJCHMAN, JOHN, and CORNEL WEST, eds.: *Post-analytic Philosophy*. New York: Columbia, 1985. xxx, 275 p. Cloth $35, paper $14.50.

RESCHER, NICHOLAS, ed.: *The Heritage of Logical Positivism*. Lanham, Md.: University Press of America, 1985. 167 p. Cloth $23.50, paper $11.25.

ROEMER, JOHN, ed.: *Analytical Marxism*. New York: Cambridge, 1986. viii, 315 p. Cloth $39.50, paper $11.95.

RORTY, RICHARD, J. B. SCHNEEWIND, and QUENTIN SKINNER, eds.: *Philosophy in History*. New York: Cambridge, 1985. xii, 403 p. Cloth $39.50, paper $9.95.

SCHOFIELD, MALCOLM, and GISELA STRIKER: *The Norms of Nature: Studies in Hellenistic Ethics*. New York: Cambridge, 1986. 287 p. $49.50.

SEEBASS, GOTTFRIED, and RAIMO TUOMELA, eds.: *Social Action*. Boston: D. Reidel, 1985. xx, 302 p. $48.

SHAFTO, MICHAEL G., ed.: *How We Know: Nobel Conference* xx. San Francisco: Harper & Row, 1986. xix, 171 p. $14.95.

SHANKER, S. G., ed.: *Philosophy in Britain Today*. Albany: SUNY Press, 1986. 315 p. Cloth $44.50, paper $18.95.

SHELP, EARL E., ed.: *Virtue and Medicine: Explorations in the Character of Medicine*. Boston: D. Reidel, 1985. xx, 364 p. $48.

SKINNER, QUENTIN, ed.: *The Return of Grand Theory in the Human Sciences*. New York: Cambridge, 1985. viii, 215 p. Cloth $29.50, paper $8.95.

SRZEDNICKI, JAN T. J., and V. F. RICKEY, eds.: *Lesniewski's Systems: Ontology and Mereology*. Boston: Nijhoff, 1986. 262 p. $49.90.

STILLMAN, PETER G., ed.: *Hegel's Philosophy of Spirit*. Albany: SUNY Press, 1986. xii, 223 p. Cloth $39.50, paper $12.95.

TAYLOR, MARK C., ed.: *Deconstruction in Context: Literature and Philosophy*. Chicago: University Press, 1986. viii, 446 p. Cloth $45, paper $16.95.

TOMBERLIN, JAMES E., ed.: *Hector-Neri Castañeda*. Boston: D. Reidel, 1986. xiii, 443 p. $69.

TRAVIS, CHARLES, ed.: *Meaning and Interpretation*. New York: Basil Blackwell, 1986. 308 p. $49.95.

UGRINSKY, ALEXEJ, ed.: *Lessing and the Enlightenment*. Westport, Conn.: Greenwood Press, 1986. vi, 188 p. $29.95.

VELASQUEZ, MANUEL, and CYNTHIA ROSTANKOWSKI, eds.: *Ethics: Theory and Practice*. Englewood Cliffs, N.J.: Prentice-Hall, 1985. xi, 499 p. $18.95.

VERMAZEN, BRUCE, and MERRILL B. HINTIKKA: *Essays on Davidson: Actions and Events*. New York: Oxford, 1985. 257 p. $24.95.

WHITE, MERRY I., and SUSAN POLLACK, eds.: *The Cultural Transition: Human Experience and Social Transformation in the Third World and Japan*. New York: Methuen, 1986. xiii, 302 p. $33.

YOVEL, YIRMIYAHU, ed.: *Nietzsche as Affirmative Thinker*. Hingham, Mass.: Nijhoff, 1986. xi, 232 p. $43.50.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NEW BOOKS

ABOULAFIA, MITCHELL: *The Mediating Self: Mead, Sartre, and Self-Determination.* New Haven, Conn.: Yale, 1986. xvii, 139 p. $16.50.

AKHUNDOV, MURAD D.: *Conceptions of Space and Time.* Cambridge, Mass.: MIT Press, 1986. 202 p. $25.

ALLAN, GEORGE: *The Importances of the Past.* Albany: SUNY Press, 1986. xiii, 260 p. Cloth $44.50, paper $14.95.

AMBROSE, ALICE, and MORRIS LAZEROWITZ: *Necessity and Language.* New York: St. Martin's, 1985. 254 p. $32.50.

ANDERSON, R. J., JOHN A. HUGHES, and W. W. SHARROCK: *Philosophy and the Human Sciences.* Totowa, N.J.: Barnes & Noble, 1986. 274 p. $28.95.

APPIAH, ANTHONY: *Assertion and Conditionals.* New York: Cambridge, 1985. xiii, 265 p. $32.50.

ARKES, HADLEY: *First Things: An Inquiry into the First Principles of Morals and Justice.* Princeton, N.J.: University Press, 1986. xii, 432 p. Cloth $50, paper $11.95.

ATWELL, JOHN E.: *Ends and Principles in Kant's Moral Thought.* Boston: Nijhoff, 1986. xiii, 226 p. $52.

AUSTIN, SCOTT: *Parmenides: Being, Bounds, and Logic.* New Haven, Conn.: Yale, 1986. xi, 203 p. $20.

AVAKIAN, BOB: *Democracy: Can't We Do Better than That?* Chicago: Banner Press, 1986. 277 p. Cloth $29.95, paper $10.95.

BAKER, A. J.: *Australian Realism: The Systematic Philosophy of John Anderson.* New York: Cambridge, 1986. xvii, 150 p. $29.95.

BAKER, G. P., and P. M. S. HACKER: *Wittgenstein: Rules, Grammar and Necessity,* vol. 2. New York: Basil Blackwell, 1986. xvi, 352 p. $49.95.

BALDWIN, LOUIS: *Portraits of God.* Jefferson, N.C.: McFarland, 1986. ix, 182 p. $18.95.

BARION, JAKOB: *Grundlinien philosophischer Staatstheorie.* Bonn: Bouvier, 1986. 146 p. 29.80 DM.

BARKER, JEFFREY H.: *Individualism and Community: The State in Marx and Early Anarchism.* Westport, Conn.: Greenwood Press, 1986. xiv, 235 p. $35.

BAUM, MANFRED: *Deduktion und Beweis in Kants transzendental Philosophie.* Königstein, Germany: Hain/Athenaum, 1986. 228 p. 68 DM.

BAUM, MANFRED: *Die Entstehung der hegelschen Dialetik.* Bonn: Bouvier, 1986. viii, 268 p. 84 DM.

BECKER, LAWRENCE C.: *Reciprocity.* London: Routledge & Kegan Paul, 1986. 436 p. $32.50.

BENHABIB, SEYLA: *Critique, Norm, and Utopia.* New York: Columbia, 1986. xiii, 455 p. $35.

BERNSTEIN, RICHARD J.: *Philosophical Profiles: Essays in a Pragmatic Mode.* Philadelphia: Pennsylvania UP, 1986. xii, 313 p. Cloth $25, paper $10.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BLUHM, WILLIAM T.: *Force or Freedom?* New Haven, Conn.: Yale, 1984. xvii, 315 p. $26.

BOËR, STEPHEN E., and WILLIAM G. LYCAN: *Knowing Who*. Cambridge, Mass.: MIT Press, 1986. xii, 212 p. $22.50.

BORMAN, WILLIAM: *Gandhi and Non-violence*. Albany: SUNY Press, 1986. xvi, 287 p. Cloth $39.50, paper $12.95.

BOSTOCK, DAVID: *Plato's* Phaedo. New York: Oxford, 1986. vii, 225 p. Cloth $39, paper $18.95.

BRAUDE, STEPHEN E.: *The Limits of Influence: Psychokinesis and the Philosophy of Science*. New York: Methuen, 1986. xiv, 311 p. $39.95.

BRECHER, ROBERT: *Anselm's Argument: The Logic of Divine Existence*. Brookfield, Vt.: Gower, 1985. viii, 138 p. $34.95.

BRENNAN, GEOFFREY, and JAMES M. BUCHANAN: *The Reason of Rules: Constitutional Political Economy*. New York: Cambridge, 1986. xiv, 153 p. $34.50.

BUDD, MALCOLM: *Music and the Emotions: The Philosophical Theories*. London: Routledge & Kegan Paul, 1985. xiii, 190 p. $24.95.

BUDZISZEWSKI, J.: *The Resurrection of Nature: Political Theory and the Human Character*. Ithaca, N.Y.: Cornell, 1986. 218 p. $22.50.

BUNGE, MARIO. *Treatise on Basic Philosophy,* vol. 7. Boston: D. Reidel, 1985. Part I: xi, 263 p. $39; Part II: xi, 341 p. $49.

BURKE, PETER: *Vico*. New York: Oxford, 1985. 101 p. $13.95.

BURRELL, DAVID B.: *Knowing the Unknowable God*. Notre Dame, Ind.: University Press, 1986. x, 160 p. $15.95.

BURROW, J. A.: *The Ages of Man: A Study in Medieval Writing and Thought*. New York: Oxford, 1986. ix, 209 p. $39.95.

CAHN, STEVEN M.: *Saints and Scamps: Ethics in Academia*. Totowa, N.J.: Rowman & Littlefield, 1986. xii, 112 p. Cloth $15.95, paper $7.95.

CARRUTHERS, PETER: *Introducing Persons: Theories and Arguments in the Philosophy of Mind*. Albany: SUNY Press, 1986. 265 p. Cloth $34.50, paper $10.95.

CARSE, JAMES P.: *Finite and Infinite Games: A Vision of Life as Play and Possibility*. New York: Free Press, 1986. 152 p. $13.95.

CHADWICK, HENRY: *Augustine*. New York: Oxford, 1986. 122 p. $14.95.

CHERNIAK, CHRISTOPHER: *Minimal Rationality*. Cambridge, Mass.: MIT Press, 1986. x, 161 p. $19.95.

CHOMSKY, NOAM: *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger, 1986. xxix, 307 p. Cloth $29.95, paper $9.95.

CHURCHILL, ROBERT PAUL: *Becoming Logical: An Introduction to Logic*. New York: St. Martin's, 1986. xviii, 597 p. $27.95.

CHURCHLAND, PATRICIA SMITH: *Neuro-philosophy: Toward a Unified Science of the Mind/Brain*. Cambridge, Mass.: MIT Press, 1986. xi, 546 p. $27.50.

CLARK, STEPHEN R. L.: *The Mysteries of Religion*. New York: Basil Blackwell, 1986. x, 277 p. $45.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COHEN, L. JONATHAN: *The Dialogue of Reason: An Analysis of Analytical Philosophy*. New York: Oxford, 1986. x, 237 p. $45.

COLEMAN, FRANCIS X. J.: *Neither Angel nor Beast*. London: Routledge & Kegan Paul, 1986. xii, 227 p. $33.

COPLESTON, FREDERICK C.: *Philosophy in Russia from Herzen to Lenin and Berdyaev*. Notre Dame, Ind.: University Press, 1986. x, 454 p. $29.95.

CORBIN, HENRY: *Temple and Contemplation*. New York: Methuen, 1986. xii, 413 p. Cloth $49.95, paper $17.95.

COTTINGHAM, JOHN GRAHAM: *Descartes*. New York: Basil Blackwell, 1986. viii, 171 p. Cloth $45, paper $14.95.

COVAL, S. C., and J. C. SMITH: *Law and Its Presuppositions*. London: Routledge & Kegan Paul, 1986. vii, 141 p. $29.95.

CREEL, RICHARD E.: *Divine Impassibility: An Essay in Philosophical Theology*. New York: Cambridge, 1986. xi, 238 p. $39.50.

CRESSWELL, M. J.: *Adverbial Modification*. Boston: D. Reidel, 1985. vii, 230 p. $44.50.

CRIMSHAW, JEAN: *Philosophy and Feminist Thinking*. Minneapolis: Minnesota UP, 1986. viii, 280 p. Cloth $35, paper $14.95.

CROWE, MICHAEL J.: *The Extraterrestrial Life Debate 1750–1900*. New York: Cambridge, 1986. xix, 680 p. $59.50.

D'AGOSTINO, FRED: *Chomsky's System of Ideas*. New York: Oxford, 1986. xii, 226 p. $34.

DAN, JOSEPH: *Jewish Mysticism and Jewish Ethics*. Seattle: Washington UP, 1986. 148 p. $20.

DAN-COHEN, MEIR: *Rights, Persons, and Organizations: A Legal Theory for Bureaucratic Society*. Berkeley: California UP, 1986. xi, 271 p. $32.50.

DANTO, ARTHUR C.: *The Philosophical Disenfranchisement of Art*. New York: Columbia, 1986. xvi, 216 p. $25.

DAUENHAUER, BERNARD P.: *The Politics of Hope*. New York: Methuen, 1987. xxii, 258 p. $47.

DELAPORTE, FRANÇOIS: *Disease and Civilization*. Cambridge, Mass.: MIT Press, 1986. xiii, 250 p. $22.50.

DELEUZE, GILLES: *Cinema 1: The Movement-Image*. Minneapolis: Minnesota UP, 1986. xiii, 250 p. Cloth $29.50, paper $13.95.

DESMOND, WILLIAM: *Art and the Absolute: A Study of Hegel's Aesthetics*. Albany: SUNY Press, 1986. xx, 222 p. Cloth $34.50, paper $10.95.

DETIENNE, MARCEL: *The Creation of Mythology*. Chicago: University Press, 1986. xii, 179 p. Cloth $25.00, paper $10.95.

DETLEFSEN, MICHAEL: *Hilbert's Program*. Boston: D. Reidel, 1986. xvi, 186 p. $42.

DIXON, KEITH: *Freedom and Equality: The Moral Basis of Democratic Socialism*. London: Routledge & Kegan Paul, 1986. 112 p. $33.

DUFF, R. A.: *Trials and Punishments*. New York: Cambridge, 1986. ix, 320 p. $32.50.

DWORKIN, RONALD: *Law's Empire*. Cambridge, Mass.: Harvard, 1986. xiii, 470 p. $20.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ECO, UMBERTO: *Art and Beauty in the Middle Ages*. Hugh Bredin, trans. New Haven, Conn.: Yale, 1986. x, 131 p. $12.95.

ELSE, GERALD F.: *Plato and Aristotle on Poetry*. Chapel Hill: North Carolina UP, 1986. xx, 221 p. $27.

ELSTER, JON: *An Introduction to Karl Marx*. New York: Cambridge, 1986. vii, 200 p. Cloth $29.95, paper $8.95.

ENGELHARDT, H. TRISTRAM, JR.: *The Foundations of Bioethics*. New York: Oxford, 1986. xvi, 398 p. $27.95.

FADEN, RUTH R., and TOM L. BEAUCHAMP: *A. History and Theory of Informed Consent*. New York: Oxford, 1986. xv, 392 p. $29.95.

FALK, W. D.: *Ought, Reasons, and Morality*. Ithaca, N.Y.: Cornell, 1986. 291 p. $29.95.

FANG, WAN-CHUAN: *A Study of Davidsonian Events*. Nankang, Taipei: Institute of American Culture, Academia Sinica, 1985. xiv, 127 p.

FEINBERG, JOEL: *Harm to Others*. New York: Oxford, 1984. xiii, 269 p. $29.95.

FINE, ARTHUR: *The Shaky Game: Einstein, Realism, and the Quantum Theory*. Chicago: University Press, 1986. xi, 186 p. $25.

FINE, KIT: *Reasoning with Arbitrary Objects*. New York: Basil Blackwell, 1986. viii, 220 p. $29.95.

FISSE, BRENT, and PETER A. FRENCH: *Corrigible Corporations and Unruly Law*. San Antonio, Texas: Trinity UP, 1985. 233 p. Cloth $30, paper $14.95.

FLEW, ANTONY: *David Hume*. New York: Basil Blackwell, 1986. viii, 189 p. Cloth $45, paper $14.95.

FONTINELL, EUGENE: *Self, God and Immortality: A Jamesian Investigation*. Philadelphia: Temple UP, 1986. vii, 297 p. $34.95.

FORRER, RICHARD: *Theodicies in Conflict: A Dilemma in Puritan Ethics and Nineteenth-century American Literature*. Westport, Conn.: Greenwood, 1986. xiv, 288 p. $37.95.

FOSTER, JOHN: *A. J. Ayer*. London: Routledge & Kegan Paul, 1985. xiii, 307 p. $32.95.

FRANK, S. L.: *The Spiritual Foundations of Society*. Boris Jakim, trans. Athens: Ohio UP, 1986. ix, 182 p. $22.95.

FRANKLIN, ALLAN: *The Neglect of Experiment*. New York: Cambridge UP, 1986. xii, 290 p. $42.50.

FRIEDMAN, ALAN J., and CAROL C. DONLEY: *Einstein as Myth and Muse*. New York: Cambridge, 1985. xiv, 224 p. $39.50.

FRIEDMAN, MAURICE: *Martin Buber and the Eternal*. New York: Human Sciences Press, 1986. 191 p. $16.95.

FRISBY, DAVID: *Fragments of Modernity*. Cambridge, Mass.: MIT Press, 1986. viii, 319 p. $27.50.

FROST, MERVYN: *Towards a Normative Theory of International Relations*. New York: Cambridge, 1986. viii, 241 p. $37.50.

FULLER, ROBERT C.: *Americans and the Unconscious*. New York: Oxford, 1986. 248 p. $19.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

FUNKENSTEIN, AMOS: *Theology and the Scientific Imagination from the Middle Ages to the Seventeenth Century.* Princeton, N.J.: University Press, 1986. xii, 421 p. $47.50.

FYNSK, CHRISTOPHER: *Heidegger: Thought and Historicity.* Ithaca, N.Y.: Cornell, 1986. 246 p. $24.95.

GARDNER, DANIEL K.: *Neo-Confucian Reflection on the Canon.* Cambridge, Mass.: Harvard, 1986. vii, 181 p. $20.

GASCHÉ, RODOLPHE: *The Tain of the Mirror: Derrida and the Philosophy of Reflection.* Cambridge, Mass.: Harvard, 1986. x, 348 p. $25.

GAUTHIER, DAVID: *Morals by Agreement.* New York: Oxford, 1986. vi, 367 p. $39.95.

GERVAIS, KAREN GRANDSTRAND: *Redefining Death.* New Haven, Conn.: Yale, 1987. xi, 231 p. $22.50.

GILMOUR, JOHN C.: *Picturing the World.* Albany: SUNY Press, 1986. xi, 214 p. Cloth $29.50, paper $9.95.

GLOBUS, GORDON: *Dream Life, Wake Life: The Human Condition through Dreams.* Albany: SUNY Press, 1987. x, 203 p. Cloth $34.50, paper $10.95.

GLOUBERMAN, M.: *Descartes: The Probable and the Certain.* Amsterdam: Rodolpi, 1986. 374 p. $39.95.

GOCHET, PAUL: *Ascent to Truth: A Critical Examination of Quine's Philosophy.* Munich: Philosophia Verlag, 1986. 202 p. 59 DM.

GOLDMAN, ALVIN I.: *Epistemology and Cognition.* Cambridge, Mass.: Harvard, 1986. viii, 437 p. $27.50.

GOUINLOCK, JAMES: *Excellence in Public Discourse: John Stuart Mill, John Dewey, and Social Intelligence.* New York: Teacher's College Press, 1986. xiv, 173 p. $16.95.

GRAHAM, GORDON: *Politics in Its Place: A Study of Six Ideologies.* New York: Oxford, 1986. viii, 196 p. $24.95.

GRAYLING, A. C.: *Berkeley: The Central Arguments.* La Salle, Ill.: Open Court, 1986. xii, 218 p. $26.95.

GRIFFITHS, PAUL J.: *On Being Mindless: Buddhist Meditation and the Mind-Body Problem.* La Salle, Ill.: Open Court, 1986. xxii, 220 p. $12.95.

GRISWOLD, CHARLES L., JR.: *Self-Knowledge in Plato's* Phaedrus. New Haven, Conn.: Yale, 1986. xii, 314 p. $29.50.

GUSTAFSON, DONALD F.: *Intention and Agency.* Boston: D. Reidel, 1986. vi, 209 p. $39.

HALBASCH, KEITH: *Philosophy: A Tough-minded Contemporary Approach.* Buffalo, N.Y.: Prometheus, 1986. 123 p. $14.95.

HALLIWELL, STEPHEN: *Aristotle's* Poetics. Chapel Hill: North Carolina UP, 1986. xi, 369 p. $30.

HALVERSON, WILLIAM H.: *A Concise Logic.* New York: Random House, 1984. xvii, 403 p. $17.

HANDWERK, GARY J.: *Irony and Ethics in Narrative from Schlegel to Lacan.* New Haven, Conn.: Yale, 1986. ix, 231 p. $21.

HANSON, KAREN: *The Self Imagined: Philosophical Reflections on the Social Character of Psyche.* New York: Methuen, 1987. vii, 152 p. $26.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIV, NO. 8, AUGUST 1987

UTILITARIANISM AND EXPECTED UTILITY*

THE literature of economics contains a formal theorem that looks on the face of it like an argument for utilitarianism. It would be remarkable if a formal argument could establish a moral theory. This one does not; but I think it can contribute to our understanding of utilitarianism. This paper explores its contribution.

Granted some superficially plausible assumptions, the theorem shows that a utility function representing "social preferences" can be written as a sum of utility functions representing the preferences of individuals. As it stands, this theorem is uninformative, because the notion of social preferences is obscure. And I shall be showing that, when the notion is elucidated, the assumptions lose their plausibility. I shall therefore reinterpret the theorem in terms of *good* rather than *preferences*, in a way that makes the assumptions reasonably acceptable. The reinterpreted theorem shows that a utility function representing general good can be written as a sum of utility functions representing the good of individuals. This seems to suggest that good is the total of individual good, so there is no value in equality in the distribution of good. I shall be considering how far the theorem really licenses this utilitarian conclusion.

I. THE FORMAL THEOREM

Suppose there are h people. Each has preferences among a set of alternative prospects, the same set for everyone. Each person's preferences satisfy the axioms of expected-utility theory—I shall call such preferences *coherent*. Expected-utility theory tells us that coherent preferences can be *represented* by a utility function. This function assigns a utility to each prospect in such a way that, of any

* This is a revision of part of a paper I wrote while I was a Visiting Fellow at All Souls College, Oxford, supported by a grant from the Social Science Research Council. Some of the revision was done while I was a Visiting Fellow at the Australian National University. I thank these institutions, and also Martin Browning, Angus Deaton, Jan Graaf, James Mirrlees, Derek Parfit, and Amartya Sen.

0022-362X/87/8408/0405$01.80 © 1987 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

two prospects, the preferred one has the higher utility. The function will also be *expectational,* by which I mean that the utility it assigns to a prospect whose results are uncertain is the mathematical expectation of the utility it assigns to the results. If a person's preferences are coherent, there are actually many expectational utility functions that will represent them, all positive linear transforms of each other.[1]

Suppose there are also social preferences among the same set of prospects. If these too are coherent, they can be represented by an expectational utility function. Once again there are actually many expectational utility functions that will represent them, all positive linear transforms of each other.

Suppose next that the social preferences are *Paretian,* that is:

> If everyone is indifferent between some pair of prospects, then the social preferences are indifferent too, and, if at least one person prefers the first of two prospects to the second and no one prefers the second to the first, then the first is socially preferred to the second.

The subject of this paper is:

> *Theorem 1.*[2] Assume that each person's preferences are coherent, and that social preferences are coherent and Paretian. Then there are expectational utility functions $U_1, \ldots U_h$ representing the individual preferences, and an expectational utility function U_g representing social preferences, such that for any prospect P
>
> $$U_g(P) = U_1(P) + \ldots + U_h(P)$$

A difficulty in understanding Theorem 1 is that it says nothing about *which* of the many utility functions that represent a person's

[1] In Richard Jeffrey's version of expected-utility theory, *The Logic of Decision,* 2nd ed. (Chicago: University Press, 1983), a wider range of transformations is allowed. Theorems 1 and 2 below are true within Jeffrey's theory, subject to some extra assumptions (see John Broome, "Bolker-Jeffrey decision theory reveals some cracks in an axiomatic buttress of utilitarianism," Discussion Paper 175, Economics Dept., Univ. of Bristol); but in this paper I confine my attention to more restrictive theories such as Leonard Savage's, see *The Foundations of Statistics,* 2nd ed. (New York: Dover, 1972).

[2] Among the published proofs of this theorem are: Robert Deschamps and Louis Gevers, "Separability, Risk-bearing and Social Welfare Judgments," in *Aggregation and Revelation of Preferences,* Jean-Jacques Laffont, ed. (Amsterdam: North-Holland, 1979), pp. 145–160; Peter C. Fishburn, "On Harsanyi's Utilitarian Cardinal Welfare Theorem," *Theory and Decision,* XVII, 1 (July 1984): 21–28; Peter J. Hammond, "Ex-ante and Ex-post Welfare Optimality under Uncertainty," *Economica,* XLVIII (August 1981): 235–250; Peter J. Hammond, "Ex-post Optimality as a Dynamically Consistent Objective for Collective Choice under Uncertainty," in *Social Choice and Welfare,* P. K. Pattanaik and M. Salles, eds. (Amsterdam: North-Holland, 1983), pp. 175–205; and John C. Harsanyi, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy,* LXIII, 4 (August 1955): 309–321.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

preferences is the one that is added to others to make up social utility. If U_1 is a utility function for person 1, then so is $2U_1$. But, if $2U_1$ were to replace U_1 in the formula above, the weight of person 1's preferences in social preferences would be increased. Social preferences would be more inclined to favor person 1 in the distribution of goods. So the implications of Theorem 1 for distribution are not clear cut.

One conclusion about distribution can be drawn nevertheless. The theorem links people's attitudes to uncertainty with the attitude of social preferences to inequality. The shape of a person's utility function indicates her attitude to uncertainty. But, when people's functions are added up to make social utility, as Theorem 1 says they can be, their shape also indicates the attitude of social preferences to inequality.

Suppose, for instance, that each person cares only about her own wealth and is risk averse about it. Then the utility function representing each person's preferences will be a strictly concave (downwards bending) function of her wealth. The social-utility function, by Theorem 1, will be a sum of such strictly concave functions. A function of this sort tends to favor equality in the distribution of wealth. The tendency may be weak, because the function may give more weight to some people's wealth than other's. But strict concavity amounts to "diminishing marginal utility of wealth," whose egalitarian tendency, though weak, is well known. Contrast it with the case where each person is risk neutral about her wealth. Then each person's utility function will be linear in her wealth. Social utility, by Theorem 1, will be linear in each person's wealth. And a linear function definitely attaches no value to equality in the distribution of wealth.

The linking of attitudes to uncertainty and attitudes to inequality is one remarkable consequence of Theorem 1. And the theorem has also been taken as an argument for utilitarianism.[3] But, as I say, I think it needs to be reinterpreted if its assumptions are to be acceptable. The reinterpreted theorem also makes a link between uncertainty and inequality, which I shall assess. And I also assess how far it supports utilitarianism.

Section II of this paper explains why a reinterpretation is needed. Section III describes the reinterpretation. Sections IV and V assess the theorem's assumptions as reinterpreted. Sections VI and VII consider what conclusions can be drawn.

[3] See John Harsanyi in "Morality and the Theory of Rational Behaviour," *Social Research*, XLIV, 4 (Winter 1977), reprinted in Amartya Sen and Bernard Williams, eds., *Utilitarianism and Beyond* (New York: Cambridge, 1982), pp. 39–62.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

One preliminary point needs to be made. Many of the proofs of Theorem 1 allow for subjective probabilities. They use versions of expected-utility theory in which the probabilities a person attaches to different events are revealed by her preferences. Different people's preferences may reveal different probabilities attached to the same event. The proofs of Theorem 1, however, also prove[4]

> *Theorem 2.* Assume that each person's preferences are coherent, and that social preferences are coherent and Paretian. Then the social preferences and all the individual preferences must agree about the probabilities of every event.

Since it is unlikely that everyone will agree about probabilities, this theorem tells us that, as a general rule, coherent Paretian social preferences cannot exist. But, if they do, then Theorem 1 tells us that they can be represented by a utility function that is the sum of individual-utility functions.

II. THE LIBERAL ARGUMENT

Are the assumptions of Theorem 1 plausible?

To test the Paretian assumption, consider this example. There are only two people. Both are concerned only about their own wealth, and they are risk neutral about it. One of two possible events will occur. Compare these prospects

	Event 1	Event 2
Prospect *A*	(2,2)	(2,2)
Prospect *B*	(3,0)	(0,3)

The parentheses show what wealth will result for the two people from each prospect if a particular event occurs. The first person attaches probability .7 to event 1 and .3 to event 2; the second person .3 and .7. Consequently, both people prefer prospect *B* to *A*. The Paretian assumption says that for this reason *B* is socially preferred to *A*. Can this be justified?

This depends on what, exactly, is meant by 'socially preferred'. Social preference is an ambiguous notion. "*B* is socially preferred to *A*" may mean simply that *B* is better than *A*. Under this interpretation, the Paretian assumption might be defended on the grounds that *B* is better than *A* because it is better for both people. This type of argument is explored in Section III. But it is not really available here because, although both people prefer *B* to *A*, *B* cannot actually be better for both of them. It can only be better for person 1 if event 1 is much more probable than event 2. And then it is not better for person 2.

[4] See the proofs by Deschamps and Gevers and Hammond mentioned in note 2.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Alternatively, "*B* is socially preferred to *A*" may mean that, if society had a choice between *A* and *B*, it should choose *B*. I take this to mean that the social system—perhaps the economic or political system, perhaps the government—should bring about *B* rather than *A*. There is no necessary implication that *B* is better than *A*.

It can plausibly be argued that, since both people prefer *B* to *A*, *B* is socially preferred to *A* in this second sense: it should come about rather than *A*. The argument is that the role of the social system is to act as a mechanism whereby people's wants get put into effect. So far as possible, it should simply bring about what people want. Normally, people's wants conflict, and then it should act as a fair mechanism for resolving the conflict. But, when there is no conflict, it should simply do what everyone wants. This argument does not insist that *B* is better than *A*. What it values is the nature of the social system, not its outcome. It says that the social system should be such that *B* comes about rather than *A*. But the occurrence of *B* is not necessarily any benefit added to the existence of the system that brings it about. Someone who accepts this argument might also believe that *B* is actually worse than *A*.

I call this *the liberal argument*. It supports the Paretian assumption under this interpretation of social preference, and I think it is the only argument that can do so. But someone who used this argument could not then defend Theorem 1's other main assumption that social preferences are coherent. The axioms of expected-utility theory are appropriate for an agent that makes choices, and derives reasons for making them from the value it attaches to their results. But the liberal argument does not conceive social preferences as representing the choices of any agent at all.[5] The social system does not really choose; it is a mechanism through which a result emerges from the choices of many agents. Social preferences represent the results that should emerge from this mechanism. And the reasons why the results should be one thing rather than another are not derived from the value of the results themselves, but from the merits of the mechanism.

Furthermore, anyone who supports the Paretian assumption has to tolerate incoherent social preferences. Theorem 2 says that social preferences, if they are to be Paretian, will be incoherent whenever people disagree about probabilities. Indeed, preferring *B* to *A* is

[5] A complication is that the working of the social mechanism often depends on the decisions of an agent. For instance, a minister in the government may actually decide between *A* and *B* in the example. And rationality requires an agent to have consistent preferences. This complication is discussed in my "Should social preferences be consistent?" Discussion Paper 180, Economics Dept., Univ. of Bristol.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

already very near to incoherence. Any decent mechanism for resolving conflicting preferences would bring about (2,2) for sure in preference to either (3,0) for sure or (0,3) for sure, because (2,2) has more wealth in total and has it equally distributed. But it then violates the axioms of expected-utility theory[6] for it to bring about *B*, which is a prospect of either (3,0) or (0,3), in preference to *A*, which is (2,2) for sure. Theorem 2 is a snare in the path of any argument designed to justify the assumptions of Theorem 1. It says that the coherence and Paretian assumptions for social preferences are inconsistent with each other, unless people's probabilities coincide. So any argument that supports the Paretian assumption, even when people's probabilities differ, has to give up the coherence assumption as a general rule. It will then be hard for it to explain why coherence should be required in the one special case when people's probabilities happen to agree.

The liberal argument, then, cannot provide a foundation for both the Paretian and coherence assumptions of Theorem 1.

III. THE REINTERPRETATION

Now I return to the other type of argument that might be used to support the Paretian assumption. It depends on this *principle of personal good,* as I call it:

> If two prospects are equally good for everyone, they are equally good, and, if the first of two prospects is better for someone than the second, and at least as good for everyone, then it is better.

In brief: the goodness of a prospect depends only—and positively—on how good it is for people.

This principle does not actually support precisely the Paretian assumption. We have already seen that it does not support it in the example of prospects *A* and *B*. To bring it to bear on Theorem 1 we shall have to reinterpret the theorem. We shall have to replace preference relations with relations of betterness. Instead of "Person *i* prefers *P* to *Q* or is indifferent between them" we must have "*P* is at least as good for person *i* as *Q*." Instead of "*P* is socially preferred or indifferent to *Q*" we must have "*P* is at least as good as *Q*."

These betterness relations ("good relations") may satisfy the axioms of expected-utility theory. If they do—if they are coherent—they can be represented by expectational utility functions. For instance, there will be a utility function U_i such that $U_i(P)$ is greater than $U_i(Q)$ if and only if *P* is better than *Q* for person *i*.

[6] Specifically Savage's sure-thing principle (21–23).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The reinterpreted theorem[7] is:

Theorem 1′. Assume that each person's good relation is coherent, and that the general good relation is coherent and satisfies the principle of personal good. Then there are expectational utility functions $U_1, \ldots U_h$ representing individual good, and an expectational utility function U_g representing general good, such that for any prospect P

$$U_g(P) = U_1(P) + \ldots + U_h(P)$$

This theorem has more acceptable assumptions than the original does. In the rest of this section I explain the reinterpretation in more detail. Then in Sections IV and V I assess the reinterpreted assumptions.

The reinterpretation replaces social preference with general good. This is more a disambiguation than a reinterpretation. As I said earlier, 'is socially preferred to' can in one sense be taken to mean "is better than."

The reinterpretation replaces individual preference with individual good. This *is* a significant change. I am making it chiefly for the sake of generality. The argument we are discussing aims to say something about how the good of individuals is aggregated together to constitute general good. For that, there is no need to adopt any particular theory about what the good of individuals consists in. One theory is that it consists in the satisfaction of their preferences. This theory is, in a way, built into Theorem 1. But Theorem 1′ can accommodate any adequate theory of individual good.

We have seen already, however, that the preference-satisfaction theory of good is not adequate. It cannot be quite right. In the example above, *B* is preferred to *A* by both people, but it cannot be better for both. So Theorem 1′ cannot accommodate precisely this theory. But it may be able to accommodate a modified theory in which a person's good consists in the satisfaction of what her preferences would be if they were "purified" by full information and sound deliberation.

In the example, each person's preference for B may be rational: she may have judged the probabilities rationally on the information she has. So 'better for' does not mean the same as 'rationally preferred by'. 'Better for' requires that the probabilities, as well as being rationally arrived at, are correct. The principle of personal good,

[7] Peter Hammond also reinterprets the theorem in a way that is, if I understand it, similar to mine; see "On Reconciling Arrow's Theory of Social Choice with Hansanyi's Fundamental Utilitarianism," in George Feiwel, ed., *Arrow and the Foundations of the Theory of Economic Policy* (London: MacMillan, 1987), pp. 179–222.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

then, works with the correct probabilities, which are the same for everyone, rather than with the probabilities people happen to believe in. In this way it avoids the trap set by Theorem 2; the principle is consistent with assuming that the general-good relation is coherent.

That is an advantage of my reinterpretation. A disadvantage is that it has to assume the existence of correct probabilities, which many subjectivists deny. I have been saying that the term 'better for' presupposes a notion of correctness for probabilities. So these subjectivists ought not to use this term. But the important thing is that they cannot avoid the trap of Theorem 2. Whatever argument they produce for the Paretian assumption or any similar principle, it will have to allow for disagreement about probabilities. And then social preferences will normally be incoherent. The subjectivists' argument will therefore have to tolerate incoherence. So it will not be able to offer a basis for Theorem 1 or any similar theorem.

These subjectivists, then, will not be impressed by a theorem of this sort. I therefore have to part company with them. I continue on the assumption that correct probabilities exist.

IV. THE PRINCIPLE OF PERSONAL GOOD

Now I come to assess the assumptions of Theorem 1, starting with the principle of personal good.

Consider first this principle as it applies to outcomes rather than prospects. It might be thought that some outcomes are simply good, or perhaps good for society, without being good for anybody. An example particularly germane to this paper is equality in the distribution of good between people. Two different ideas need to be looked at here.

One is the idea that, for the same total of personal good, it is better for it to be more equally distributed than less. This may suggest that equality is in a sense a good over and above the good of people. Utilitarianism denies this idea, and Theorem 1′ seems designed to support utilitarianism in this; section VII considers how far it really does so. Therefore, it had better not be simply an assumption of the theorem that the idea is wrong. And, fortunately, it is not, because the idea is consistent with the principle of personal good. Suppose, for instance, that general good is the product of individual goods—$G_g = G_1G_2 \ldots G_h$—and that all individual goods are positive. Then, for a given total of individual good, general good is increased by distributing it more equally. But, nevertheless, general good depends only—and positively—on individual good. And this is all the principle of personal good requires.

The second idea is that it is sometimes better to increase equality by taking good from people who have a lot, even if this does not

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

benefit people who have less. This *does* conflict with the principle of personal good. And I think it is wrong. If such a change is really an improvement, that can only be because it improves fairness. The less well off are more fairly treated. But then this fairer treatment is a sort of good that is done them. Either there is such a good, in which case the change benefits the less well off after all, or there is not, in which case the change is not an improvement.

To speak generally, if any ostensible improvement cannot be pinned down as good for someone, I am not inclined to believe it is really an improvement. And from now on I am simply going to take the principle of personal good for granted when applied to outcomes: for outcomes, good depends only on people's good.

But there is a special difficulty about applying it to prospects. The sense in which the notion of good applies to prospects is a derivative and attenuated one. What is really good is not the prospect but the event it is the prospect of. Suppose some good event is going to happen to me. The prospect of this event is a good prospect. But when the good in my life is catalogued the prospect will not be recorded as an extra item besides the event. Certainly, the good prospect might *cause* some good that must be recorded: enjoyable anticipation, for instance. But it is not because of such effects that the prospect is good; it would still be good even if it did not cause any. Suppose now that some good event is likely but not certain to happen to me. That is a good prospect for me. But, if unluckily the event does not happen, then my life has not been made better by the fact that I once had this good prospect. The fact that something good might have happened is not itself something good that did happen.

Applied to this attenuated sort of good, the principle of personal good encounters a difficulty. Take this example (again, the parentheses show two people's wealth):

	Event 1	Event 2
Prospect *C*	(1,1)	(1,1)
Prospect *D*	(0,2)	(2,0)

Suppose that the events are equally probable. Suppose *C* is better than *D* for both people; this will be true if it is good for both of them to avoid risk. (This may be obscure; I shall come back to it.) Then the principle of personal good says that *C* is better than *D* because it is better for both people. Consider an act of choosing *C* rather than *D*. The principle offers as a reason in favor of this act that its consequence will be better for both people. But actually the consequence will be better for one and worse for the other: one will have one unit of wealth when she would otherwise have had nought, but the other

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

will have one unit when she would otherwise have had two. The principle of personal good is trying to insert between the act of choice and its consequence a sort of quasi consequence, the prospect. And it applies to such quasi-consequences ways of thinking that may be very plausible when applied to real consequences but cannot be taken for granted here. How can it be argued that choosing *C* rather than *D* has results that are better for both people when, plainly, it does not?[8]

To deal with this point, we have to consider in detail what reasons there really are for choosing *C* rather than *D*, or *D* rather than *C*. Generally, we have to consider what reasons should guide choices between uncertain prospects. Talk about the goodness of alternatives ultimately aims to supply reasons for choice. So when the talk about goodness is in doubt we must look directly at the reasons.

To do this properly we need to have in mind a particular agent. It may be that a reason for one agent is not a reason for another, or that different agents should give different weights to the same reasons. So let us fix on an impartial agent with a general duty to act rightly but no special duty to any particular person. Perhaps the government might suit this role.

What reasons might such an agent have for choosing one alternative rather than another? There are, first of all, reasons of the sort that underlie the liberal argument described in section II. Suppose everyone prefers one of two alternatives to the other. Democratic principles suggest that this is the one that should come about, whether or not it is better for everybody, or for anybody. As I said, this is an argument about how the social system should operate as a mechanism. It is not necessarily a reason for any particular agent to bring about this alternative. But our agent may be, like the government, part of the social system. Its decisions may indeed determine what the social system brings about. In that case, these democratic considerations will constitute a reason for the agent to act in a particular way.

But we are not now concerned with reasons of this sort. We are concerned with good, and they are specifically not directed at good. I explained in section II that reasons like these are likely to lead to incoherent preferences, and that is why I turned to consider good instead. So from now on, I shall leave them aside and consider only good-directed reasons.

[8] I asked this, too forcefully, as a rhetorical question in my "Trying to Value a Life," *Journal of Public Economics*, IX (1978): 91–100. There is actually a good answer to it, which I am about to give. But most of that paper is, I still think, correct.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

By a "good-directed" reason I mean a reason that is directed toward good in the outcome of a choice. The notion of a good-directed reason, then, does not depend on the notion of good applied to prospects. So I can use it to define good applied to prospects. When I say that one prospect is better than another, I mean that an impartial agent has stronger good-directed reasons for bringing about the first (if it can) than the second. And when I say that one prospect is better for a person than another, I mean that there are stronger reasons directed toward this person's good for bringing about the first than the second.

This way of defining good for prospects meets one essential requirement. Suppose we compare two prospects that each lead for sure to a particular outcome. The better prospect, as I have defined it, is the one that leads to the better outcome. Theorem 1′, which is formally about the goodness of prospects, will therefore tell us something about the goodness of outcomes too.

With these definitions, it is easy to produce an argument for the principle of personal good applied to prospects. Suppose two prospects are equally good for everybody. This means that for each person the reasons directed toward her good for bringing about one are exactly as strong as the reasons directed toward her good for bringing about the other. The impartial agent should act rightly, and beneficence is at least a part of acting rightly. So these will all be reasons for the agent too. And, since the reasons directed toward each person's good are evenly balanced, the agent's reasons directed toward everybody's good taken together will be evenly balanced too. As I say, I am taking it for granted that, for outcomes, good depends only on people's good. So these are all the good-directed reasons there are. Therefore, all the agent's good-directed reasons are evenly balanced. That is to say, the two alternatives are equally good. This proves the first part of the principle of personal good. The second part can be proved in a similar way.

I think this argument comes to the right conclusion. But it skips over what is really the main point. We need to consider the weighting of reasons in more detail.

Take the example of prospects *C* and *D*. Person 1's reasons in favor of *C* boil down to this: if event 1 comes about, she will get one unit of wealth instead of nought. Her reasons in favor of *D* boil down to: if event 2 comes about, she will get two units of wealth instead of one. We assumed that *C* is actually better for her. The former reason, that is to say, outweighs the latter. This is because we assumed that it is good for her to avoid risk. The value of risk avoidance appears in expected-utility theory in the form of weights attached to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

gains and losses of wealth. Formally it appears in a strictly concave utility function. Here it says that the difference between one unit and nought units has more weight than the difference between two units and one unit. Consequently, the reason in favor of *C* outweighs the reason in favor of *D*.

Our impartial agent has four good-directed reasons to weigh up. Two of them, directed toward person 1's good, are the ones I have just mentioned. And there are two symmetrical ones directed toward person 2's good. How should these reasons be weighed up in determining what the agent ought to do? We do not have to worry here about weighing against each other reasons directed toward different people's good. But I do want to argue that the two reasons directed toward person 1's good must be given the same relative weight as they receive when determining what is best for person 1.

The relative weight given to these reasons in the latter application is the means by which expected-utility theory takes account of whether or not it is good for person 1 to avoid risk, and the degree to which it is good. In our example we are assuming that risk avoidance is indeed good for person 1. So, in determining which prospect is better for her, the reason in favor of *C* outweighs the reason in favor of *D*. If, in determining which alternative our agent ought to choose, we did not give these reasons the same relative weight, the fact that avoiding risk is good for person 1 would not be properly taken account of in our determination. In the example, one reason for the agent to choose prospect *C* rather than *D* is that *C* is less risky for person 1, and avoiding risk is good for person 1. In expected-utility theory this reason appears, not strictly as a separate reason on its own, but in the weighting of other reasons. So this weighting must be preserved.

This is the point of treating prospects as "quasi consequences" and applying the notion of good to them. There is a type of reason for making a choice, namely avoiding risk,[9] which does not appear in the value of true consequences. But it can be represented in the value of prospects.

V. COHERENCE

Theorem 1′ requires individual and general good to satisfy the axioms of expected-utility theory. I mention here only the ones that are likely to give trouble: the completeness axiom and the various consistency axioms. The completeness axiom says that, of two prospects, either one is better than the other or else they are equally

[9] If, alternatively, it is good for a person to take on risk or to be neutral about risk, that too will be a reason for choice.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

good. I do not detail the consistency axioms, except to mention that the most controversial is the axiom of strong independence.

I see no reason to think that either the individual- or the general-good relations are complete. For an individual, it is plausible that there are goods that cannot be weighed against each other. For general good, there is the added difficulty that it may not be possible to weigh one person's good against another's. For the sake of Theorem 1′ we shall simply have to assume there are no such intrapersonal or interpersonal incommensurabilities. This is a weakness of the theorem.[10]

On the other hand, I believe there are good arguments to show that the individual- and general-good relations satisfy the consistency axioms. I have argued in another paper[11] that the preferences of a rational agent will satisfy these axioms. It follows that each individual-good relation will satisfy them, because a person's good relation is what her preference relation would be were she rational and self-interested, and were her probabilities correct.[12] Similarly, the general-good relation would be the preference relation of a rational impartial agent whose preferences were determined by the weighing up of good-directed reasons in the manner described in section IV. So this relation, too, must be consistent.[13]

VI. UNCERTAINTY AND INEQUALITY

I have done what I can to justify the assumptions of the reinterpreted Theorem 1′. One source of doubt about them is that they presuppose a notion of correctness for probabilities. Another is that they ignore any problems there might be about the commensurability of goods. But apart from these doubts, I think the assumptions are acceptable.

[10] Richard Jeffrey, in "On interpersonal utility theory," this JOURNAL, LXVII, 20 (October 21, 1971): 647–656, uses Theorem 1 to support the possibility of interpersonal comparisons. I think this is wrong.

[11] John Broome, "Rationality and the Sure-thing Principle," in *Rationality, Self-Interest and Benevolence,* Gay Meeks, ed. (New York: Cambridge, forthcoming).

[12] This does not mean that a person's good consists in the satisfaction of her preferences, a theory about the nature of good mentioned in section III. The determination is in the opposite direction.

[13] The argument of this paragraph is slightly too quick. There might be no such thing as rational self-interested preferences, because it might not be rational to be self-interested. Similarly, it might not be rational for an impartial agent to determine its preferences by the weighing up of good-directed reasons, because, as I explained in section IV, the liberal argument of section II may also supply reasons to such an agent, and these are not good directed. Nevertheless, if a person's preferences *were* determined by her own interests only, and if an impartial agent's preferences *were* determined by good-directed reasons only, then these preferences should satisfy the consistency axioms. It is easy to check that my arguments in "Rationality and the Sure-thing Principle" would apply to such preferences.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

So, subject to these qualifications, the conclusion of Theorem 1′ follows. Individual and general good can be represented by utility functions in such a way that general utility is the sum of the individual utilities. What does this tell us? In section I, I distinguished two conclusions one might draw from Theorem 1. First, the theorem made a link between people's attitudes to risk and the attitude of social preferences to inequality. Second, the theorem seemed to be an argument for utilitarianism. We must make a similar distinction for the reinterpreted Theorem 1′.

The first conclusion to be drawn from Theorem 1′ is that the value of equality is linked with the attitude people should take to uncertainty. If people should be neutral about risks to their wealth, then it is wrong to favor equality in the distribution of wealth. And, if people should be risk averse about wealth, there is some presumption in favor of equality in wealth. This link is explained in section I. The only difference is that now we are dealing with what is good for people faced with uncertainty, rather than with the attitude they actually take. Notice, though, that what is good for people here need not be determined by anything other than the people's own tastes. A person's tastes help to determine what is good for her.[14] If she likes apples more than pears then, other things being equal, an apple is better for her than a pear. Shifting attention from preferences to good does not deny the importance of tastes.

That inequality should be linked to uncertainty in this way strikes me as a remarkable consequence of Theorem 1′.

VII. UTILITARIANISM

Does Theorem 1′ give any more general support to utilitarianism?

At most it can support only a part of it. It has nothing to say about the utilitarian thesis that one should act so as to bring about the best result; it is only about what result is best. It has nothing to say about the utilitarian thesis that a person's good consists in pleasure or the satisfaction of her desires; it is only about how different people's good is aggregated. And it simply assumes the utilitarian thesis that goods are always commensurable.

The one part of utilitarianism that the theorem does seem to support is the thesis that general good is the sum of people's good, so there is no value in an equal distribution of good. But so far we are not entitled to suppose it says even this. It says that general utility is

[14] I distinguish tastes from preferences. I think that, other things being equal, satisfying a person's tastes is necessarily good for her. But I do not think the same about her preferences. A taste can supply a reason for a preference, but a preference can also be based on other reasons, or on no reason.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the sum of individual utilities. But utilities have been defined only to represent the *order* of good: of two prospects the one with the higher utility is the better. They do not necessarily represent *degrees* of good.

A case can be made out nevertheless for saying that utilities do actually represent degrees of good. Theorem 1′ itself can contribute to this case. But first let us see what argument can be made independently of Theorem 1′.

If a utility function represents the order of a person's good, then so does any linear transform of it. So the most that can be expected of a person's utility function in general is that it should be a linear transform of her good. The characteristic of a linear transformation is that it preserves the order of differences. Take four prospects M, N, P, and Q, each good for a person to some degree. Write these degrees $G(M)$, $G(N)$, $G(P)$, and $G(Q)$. Then the most that can be expected in general from a utility function for the person is that the utility difference $[U(M) - U(N)]$ should be greater than $[U(P) - U(Q)]$ if and only if $[G(M) - G(N)]$ is greater than $[G(P) - G(Q)]$.

Do utilities represent degrees of good to this extent? Suppose a person is faced with a choice between getting one unit of wealth for sure or alternatively taking a gamble at equal odds of either no units or two units. Suppose that

$$U(2) - U(1) < U(1) - U(0)$$

The latter difference in utility outweighs the former, and the best choice is the one unit for sure. Is it then necessarily the case that

$$G(2) - G(1) < G(1) - G(0) \qquad ?$$

This will be so if it is best for the person to maximize the expectation of her good.[15] Since utility is defined so that it is best for her to maximize the expectation of her utility, utility will then represent degrees of good to the required extent. But it is quite plausible that it is not best for a person to maximize the expectation of her good. For instance, the notion of degrees of good may not even make sense; good may not be an arithmetic quantity. Or it might be good for a person to be risk averse about good.[16] Then her utility will be a

[15] Many decision theorists seem to have thought people should be expected-good maximizers. Daniel Bernoulli, "Exposition of a New Theory on the Measurement of Risk," Louise Sommer, trans. *Econometrica*, XXII, 1 (January 1954): 23–36, thought people should maximize expected *emolumentum*, which the dictionary translates as "benefit or advantage." Jeffrey, *The Logic of Decision*, thinks they should maximize expected "desirability."

[16] Such a view is implicit in Kenneth J. Arrow, *Social Choice and Individual Values*, 2nd ed. (New Haven, Conn.: Yale, 1963), p. 10.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

strictly concave function of her good, so we might consistently suppose in the example that actually, say,

$$G(2) - G(1) = G(1) - G(0)$$

But there is a possible retort to this. It might be said that it is precisely in comparisons of the sort we are making that the notion of degrees of good gets its meaning.[17] In making decisions in the face of uncertainty, gains and losses are weighed against each other. In the example, the possible gain in wealth from one unit to two is a reason in favor of taking the gamble. The possible loss from one unit to none is a reason against. The latter is the stronger reason. This is naturally expressed by saying that the difference between one unit and no units of wealth amounts to a greater difference in good than the difference between one unit and two. According to the supposition in the previous paragraph, these differences in good are actually the same. But what can this mean, if it is not that they are evenly balanced when weighed against each other as reasons for choosing? According to the supposition, the differences are equal but they do not *count* equally as reasons. But maintaining a distinction between amounts of good and how these amounts count looks like an empty gesture.

I think this is a good retort. It is hard to see what use we can have for the notion of degrees of good except when weighing up differences in good as reasons for making a choice. So it is in weighing up differences that we can expect the notion to get its meaning. Decision making under uncertainty, however, is not the only context in which differences in good are weighed against each other. Perhaps the notion gets its meaning elsewhere. Another context where differences of good are weighed is in the distribution of good between people. Let us consider that.

Suppose there is a choice between the distributions of wealth (1,1) and (0,2). A reason in favor of the first is that it gives person 1 one unit of wealth instead of none. A reason in favor of the second is that it gives person 2 two units instead of one. How should these reasons be weighed against each other? This is what Theorem 1′ is about. (Remember that it is simply an assumption of the theorem that reasons like these *can* be weighed against each other; interpersonal comparisons are possible.) It says that we can find the right weights from the people's utility functions. To do so, we have to make sure that for each person we have picked the appropriate utility function.

[17] Compare J. A. Mirrlees, "The Economic Uses of Utilitarianism," in Sen and Williams, pp. 63–84.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Each person has many functions representing the order of her good, and the theorem gives no guidance about which is the right one. But it does say that there is a right one. And once we have it, differences in utility determine the weights that should be given to opposing reasons. In the example, once we have functions U_1 and U_2 for the people, we compare $[U_1(1) - U_1(0)]$ with $[U_2(2) - U_2(1)]$. If the former is greater, the distribution (1,1) is better; if the latter, (0,2). This is a context where we are weighing up reasons. So according to what I said above, it gives us grounds for saying that these differences of utility represent differences in degrees of good. Suppose $[U_1(1) - U_1(0)]$ is greater than $[U_2(2) - U_2(1)]$. Then we have grounds for saying that person 1 gains more in good from having one unit of wealth instead of none than person 2 gains from having two units instead of one. Since the same utility functions supply the right weights in any distributional comparison, we have grounds for saying that these functions represent degrees of good.

But these grounds are unlikely to convince a nonutilitarian. They beg the question. They insist that, when weighing reasons, the stronger reason must always be the one that represents the greater difference in good. This simply assumes that the better alternative is always the one with the greater total of good. And that was what had to be proved.

The strength of the utilitarian case, however, is this. The functions U_1 and U_2, which supply the weights when weighing reasons in distributing wealth, are utility functions for the people. Therefore, they also supply the weights when weighing reasons in making decisions under uncertainty. So these functions serve the same purpose in two contexts. This very much strengthens the claim that they represent degrees of good. This is the effect of Theorem 1′. Theorem 1′ provides a strong case for saying that utilities represent degrees of good.[18] And, having done so, it also says that the better of two alternatives is always the one with the greater total of good.

The answer to the nonutilitarian's objection is this. The objection relies on a distinction between degrees of good and how these degrees count in weighing reasons: utility tells us how good counts, but utility may be distinct from good itself. But we have been shown no way of assigning meaning to degrees of good apart from how they count. And, without that, the distinction now seems emptier than ever.

[18] Compare John C. Harsanyi, "Nonlinear Social Welfare Functions: A Rejoinder to Professor Sen," in *Foundational Problems in the Social Sciences*, R. Butts and J. Hintikka, eds. (Boston: Reidel, 1977), pp. 293–296.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

But this is not the end of the argument. All the nonutilitarian has to do now is supply a suitable way of assigning meaning to degrees of good. What she needs is *another* context in which differences of good are weighed against each other. The one to turn to, I think, is the weighing up of good at different periods of a person's life. I intend to pursue this in another paper.

Imagine for a moment, though, that the argument had come to a decisive end in the defeat of the nonutilitarian. Imagine we had managed to derive from Theorem 1′ the utilitarian conclusion that good is the sum of people's good. What would this conclusion now amount to? The way we would have come to it shows it is less significant than it seems. It seems anti-egalitarian. But our argument was simply about meaning. In order to make sense of the question whether or not there is value in an equal distribution of good, we have to assign a meaning to the notion of degrees of good. And it happens that the most natural way of doing that prevents us from attaching value to equality in the distribution of good. That is all. It suggests that the question whether there is value in equality in the distribution of good is unimportant.

Furthermore, the argument also shows that the utilitarian conclusion adds nothing at all to the conclusion we reached in section VIII. There we made a connection between uncertainty and inequality. The same utility functions that represent what is best in the face of uncertainty also represent what is best in distribution between people. And it is simply because the same functions appear in both contexts that we have now decided they represent degrees of good. Theorem 1′ gives no more general support to utilitarianism than that.

The argument, to summarize, is not ended. But, if it were, it would have achieved less than might have been expected of it.

JOHN BROOME

University of Bristol

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMPATIBILISM AND THE ARGUMENT FROM UNAVOIDABILITY*

IN recent years, a certain type of argument against compatibilism has become prevalent.[1] According to this argument, determinism is demonstrably incompatible with human freedom because determinism entails that our actions result necessarily from laws of nature and past states of the world which are themselves necessary. Since that which results necessarily from something necessary is itself necessary, it follows (given determinism) that our actions are also necessary and, hence, not free.

Though widely discussed, this type of argument has not yet acquired an accepted name. Since the notion of necessity which plays so crucial a role in it seems to be equally well expressed by such terms as 'inevitability', 'uncontrollability', 'inescapability', 'unalterability', or 'unavoidability', any one of these might provide us with a suitable name. Since I shall (for no compelling reason) phrase the argument in terms of unavoidability, I shall christen it the *Argument from Unavoidability;* for purposes of brevity, though, I shall often condense this simply to: the *Argument.*

In section I of this paper, I set forth the Argument from Unavoidability in some detail and show that its success depends upon the incompatibilist's ability to fashion a notion of unavoidability which meets three separate conditions. In section II, I examine various putative analyses of unavoidability to see whether any of them does satisfy our three conditions. My conclusion is that no candidate is likely to pass all three tests unless a certain thesis regarding the prerequisites of power can safely be assumed. I argue in section III, however, that such a thesis cannot safely be assumed in the context of a purported proof of incompatibilism. Accordingly, the Argument

* This research was supported by a grant from the American Council of Learned Societies funded through the National Endowment for the Humanities.

[1] The argument is perhaps most commonly associated with the name of Peter van Inwagen; see his *An Essay on Free Will* (New York: Oxford, 1983), ch. III; as well as his earlier "The Incompatibility of Free Will and Determinism," *Philosophical Studies*, XXVII, 3 (March 1975): 185–199. Other versions of the argument are offered by Carl Ginet, "Might We Have No Choice?" in Keith Lehrer, ed., *Freedom and Determinism* (New York: Random House, 1966), pp. 87–104; James Lamb, "On a Proof of Incompatibilism," *Philosophical Review*, LXXXVI, 1 (January 1977): 20–35; and David Wiggins, "Towards a Reasonable Libertarianism," in Ted Honderich, ed., *Essays on Freedom of Action* (London: Routledge & Kegan Paul, 1973), pp. 31–61. Though the argument has received much attention in recent years, it would probably be rash to acclaim it a "new" argument; see Wiggins, p. 59, note 18.

0022-362X/87/8408/0423$01.80 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

fails. I conclude by briefly commenting on the implications of this failure for the over-all dispute.

I

Determinism is the thesis that the state of the universe at any time is the logical consequence of the state of the universe at any earlier time conjoined with the laws of nature.[2] According to incompatibilists, the truth of this thesis would mean that the belief in human freedom is a delusion; freedom can exist only if the world is *not* deterministic. The Argument from Unavoidability attempts to prove that indeterminism is indeed a necessary condition of human freedom, by showing that any action performed in a deterministic universe would be necessary, or unavoidable, for the agent and, hence, could not rank as a free action.

The heart of the Argument has been nicely summarized by one of its chief critics, Michael Slote[3]:

> Given determinism, there will always be some such earlier set of conditions *s* that is connected by laws of nature to any given human action *a* that takes place. But nothing can be done to alter, nothing can be *done* about those laws; and neither, it may be added, can anything be done about *s* at any time when the doing of *a* is immediately in question. Since *s* is thus necessary (e.g., unalterable) at the time when *a* is in question and since a law leading from *s* to *a* is similarly necessary (unalterable) at that time, it would seem to follow that *a* itself is necessary (unalterable, unavoidable) at that time—however ignorant of that fact the agent of *a* may be. Presumably, then, the agent in question does not act freely in performing *a*, and since the argument has been entirely general in its assumptions, one may conclude that no human being ever acts freely in a deterministic universe (6).

Let us try to specify the Argument a bit more formally. Take some everyday action that the compatibilist would want to label free—say, Elfreth's petting his iguana at time *t*. Let '*A*' stand for the proposition that Elfreth pets his iguana at *t*. If determinism is true, *A* is the necessary consequence of any earlier state of the universe; in a deterministic world, the laws of nature—laws which we humans are powerless to break—guarantee that any such previous state ulti-

[2] I am well aware that the definition of determinism offered here might have to be modified to account for various objections—e.g., that it leaves no room for the occurrence of miracles in a deterministic universe. As I am convinced that no such modification would make any difference to our discussion of the Argument and as things are going to get quite complicated enough in the course of that discussion, I have decided to go with the simpler definition.

[3] "Selective Necessity and the Free-will Problem," this JOURNAL, LXXIX, 1 (January 1982): 5–24.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

mately leads to that state of the world at t which includes Elfreth's iguana petting. Thus, if we use 'N' to express the relevant type of necessity and 'S' to stand for a proposition expressing some such prior world-state, we can say that determinism is true only if

(1) $$N(S \supset A)$$

But, at the time t when Elfreth pets his iguana, there is nothing whatsoever that Elfreth can do about the truth of S, for humans simply have no power over prior states of the universe. Thus, at t, S was also necessary or unavoidable for Elfreth—that is,

(2) $$N(S)$$

Now, (1) states that a certain conditional is necessary, and (2) tells us that the antecedent of this conditional is also necessary. From this it surely seems to follow that the consequent is necessary as well—i.e., that

(3) $$N(A)$$

is also true. But to say that A is necessary is to say that it is unavoidable for Elfreth—that there is nothing he can do to keep it from being true. But, if Elfreth is thus powerless with regard to the truth value of A, it seems clear that his action is *not* a free action.[4] Thus, if determinism is true, Elfreth's action is not free. And, since performing free actions is no more up Elfreth's alley than it is up that of any agent, the upshot of the Argument is that determinism is indeed incompatible with human freedom. To uphold freedom—i.e., to reject (3)—we need to reject either (1) or (2); and, since (2) is clearly true, those who wish to say that humans are free have no choice but to reject (1) and the deterministic thesis from which (1) follows.

The obvious first step for compatibilists seeking to defuse this line of reasoning would be to focus on the notion of necessity employed in the Argument. The use of a necessity operator is, of course, what lends the Argument considerable strength; but is this strength ac-

[4] This conditional may not be an instance of a general truth; there may be cases where powerlessness over the performance of an action does *not* entail that the action is unfree. See Harry Frankfurt, "Alternate Possibilities and Moral Responsibility," this JOURNAL, LXVI, 23 (Dec. 4, 1969): 829–839. Frankfurt's thesis is, of course, a controversial one, but the correct resolution of that controversy is of little or no relevance to the Argument. For Frankfurt's examples cast doubt at most on the principle, "If X performs Y freely, X could have refrained from performing Y"; they give us no reason at all to doubt the truth of the principle, "If X performs Y freely, X could have refrained from performing Y freely." But the latter principle is all we need for the Argument, provided that we amend A to read, "Elfreth pets his iguana freely."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

quired legitimately? After all, the main reason we are inclined to think that (3) *does* follow from (1) and (2) is that an inference of the form

[A]
$$\begin{array}{c} N(X \supset Y) \\ N(X) \\ \hline N(Y) \end{array}$$

is evidently valid, provided that the 'N' represents *logical* or *metaphysical* necessity. But the 'N' in the Argument *cannot* be interpreted in such a way—if it were, there would be no reason at all to think that either (1) or (2) is true, and hence the Argument would collapse. Rather, the notion of necessity employed in the Argument is, as was noted, that of unavoidability. *Might* it not be, then, that the Argument's prima facie force derives in large part from the conflation of the concepts of metaphysical necessity and unavoidability?

The easiest way for the incompatibilist to meet this charge of equivocation is simply to dispense with talk of necessity and to present the Argument explicitly and exclusively in terms of unavoidability. Hence, using 'U' to stand for unavoidability, we can simply amend (1)–(3) as follows:

(1) $U(S \supset A)$

(2) $U(S)$

and

(3) $U(A)$

and thereby escape the possibility of confusion.

This escape, however, is not without cost. For now, the compatibilist can respond, how can we evaluate the Argument until the incompatibilist tells us just how this concept of unavoidability is to be understood? So long as the Argument was stated in terms of necessity, the incompatibilist could point to inference form [A] as warranting the Argument's claim to validity. But now that talk of necessity has been abandoned, the relevant rule of inference is not [A] but

[B]
$$\begin{array}{c} U(X \supset Y) \\ U(X) \\ \hline U(Y) \end{array}$$

And why should we accept [B] until we are told just what unavoidability amounts to? Indeed, in the absence of such an explication, can we even be sure that either (1) or especially (2) is true, or that the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

truth of (3) really would imply that Elfreth's action was unfree? In sum, is not the Argument's luminosity all but extinguished when we move from the (comparatively) clear land of necessity to the uncharted realm of unavoidability?

The compatibilists' response here amounts to a challenge to the incompatibilist, a challenge to provide a definition (analysis, characterization) of unavoidability such that that part of the Argument whose conclusion is (3)—a subargument henceforth to be called the *Subargument*—is clearly (i) valid, (ii) sound, and (iii) relevant to the issue of freedom. Unless such a definition can be offered, compatibilists might say, why should we think that the Argument is of any force at all against compatibilism?

It seems to me that the compatibilists' challenge here is appropriate. For the Argument to be a success, the incompatibilist needs to offer an understanding of unavoidability which satisfies the three requirements just mentioned. Let us consider each of the three in turn.

(i) *Validity*. The Argument alleges that (3) follows from (1) and (2). But this allegation is defensible only if inference form [B] is valid. Now, as Slote has noted, the most obvious way to argue for the validity of [B] is by assuming that unavoidability is both *agglomerative* (i.e., closed with respect to conjunction introduction) and *closed under entailment* (i.e., under logic implication). For, if agglomerativity holds—i.e., if

$$(\alpha) \qquad (\mathrm{U}p \,\&\, \mathrm{U}q) \supset \mathrm{U}(p \,\&\, q)$$

is true for any p and q—then we can legitimately deduce from the two premises of [B] that $\mathrm{U}[(X \supset Y) \,\&\, X]$. But, of course, $[(X \supset Y) \,\&\, X]$ entails Y. So, if unavoidability is also closed under entailment—i.e., if

$$(\beta) \qquad (\mathrm{U}p \,\&\, p \rightarrow q) \supset \mathrm{U}q$$

is also true for any p and q—then we could infer $\mathrm{U}(Y)$ from $\mathrm{U}[(X \supset Y) \,\&\, X]$, and thereby prove the validity of [B].[5] Since [B] is the only rule of inference upon which the Subargument relies, it follows that incompatibilists can guarantee the validity of the Subargument provided they make sure that their definition of unavoidability makes both (α) and (β) true.[6]

[5] Slote, pp. 9/10. My [B] corresponds to what Slote calls "the main modal principle."

[6] It is, of course, conceivable that [B] be (demonstrably) valid even though either (α) or (β) was (demonstrably) false. Slote's point is only that, given the absence of any other plausible means of justifying [B], it would be hard to feel confident of its validity if either (α) or (β) were in doubt.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(ii) *Soundness.* If the validity of the Subargument has thus been assured, its soundness will follow, provided that the definition of unavoidability employed is one that (given determinism) renders both (1) and (2) true. Of course, a definition could conceivably have *this* effect—render both (1) and (2) true—even though no one could tell (or tell for sure) that it did; but a definition that provided soundness in this minimal sense would be worthless as far as the Argument is concerned. For the Argument is supposed to provide us with a *reason* to reject compatibilism, and an argument whose soundness could not be recognized could hardly offer such a reason.[7] Hence, the Argument will succeed only to the extent that our definition of unavoidability is such that no one who presupposes the truth of determinism could reasonably reject (1) or (2).

(iii) *Relevance.* For the Argument from Unavoidability to succeed, incompatibilists must take care not to define unavoidability in such a way that the Subargument, though valid and sound, is susceptible to the charge of *ignoratio elenchi.* That is, it must be evident that the truth of (3) does indeed entail that Elfreth's action was unfree. If one can plausibly charge that the sense of unavoidability employed in the Argument is not relevant to the question of human freedom, the Argument will have failed to refute compatibilism.

For the Argument to work, then, incompatibilists need to provide an analysis of unavoidability which (i) guarantees the truth of (α) and (β), (ii) makes premises (1) and (2) clearly true if the truth of determinism is assumed, and (iii) assures us that the truth of (3) implies Elfreth's lack of freedom. What are the likely candidates for such an analysis, and how well do they satisfy these three conditions?

II

Before we look at particular candidates, a couple of general points are in order. First, since the Argument seems to view unavoidability as a property relating to propositions (rather than, say, to states of affairs or events), all the prospective characterizations of unavoidability to be examined will be attempts to define such a property; what is unavoidable, then, will be (the truth of) a proposition. Second, since it is surely conceivable that one agent could have no power over the truth of a proposition while another agent did have such power, it is conceivable that a proposition could be unavoidable for one

[7] Though Wiggins seems more cautious, there can be little doubt that Lamb and van Inwagen do (or at least once did) indeed see the Argument as *proving* or *demonstrating* that compatibilism is false. A glance at the title of Lamb's article, or at the first paragraph of ch. III of van Inwagen's *An Essay on Free Will,* should suffice to establish this point.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

agent but avoidable for another. This suggests that unavoidability is most properly seen as a relational property, with agents and propositions as relata.[8] Henceforth, then, let us read 'Uap' as "p is unavoidable for a." In accord with this new reading, we might (letting 'e' stand for 'Elfreth') reformulate (1)–(3) as

(1) $Ue(S \supset A)$

(2) $Ue(S)$

(3) $Ue(A)$

(α) and (β) would now read:

(α) $\forall a \forall p \forall q[(Uap \,\&\, Uaq) \supset Ua(p \,\&\, q)]$

(β) $\forall a \forall p \forall q[(Uap \,\&\, p \Rightarrow q) \supset Uaq]$

It is fairly easy to find analyses of unavoidability which take account of these general points and which have (at least initially) an air of plausibility, but which fail to satisfy the three conditions noted in section I. Take, for example, the notion of unavoidability favored by Slote. Noting that most proponents of the Argument would insist that the state of the universe specified by S in (1) and (2) be a state prior to Elfreth's birth, Slote speculates that their reason for so insisting is their wish to make sure that the beliefs, desires, abilities, and the like which Elfreth has at t form no part of the causal explanation of the truth of S. This insulation from the causal reach of such factors as current beliefs is, Slote insinuates, what is central to the notion of unavoidability. Hence, he suggests that we define *unavoidability* in something like the following way:

(D1) $Uap =_{df}$ The state of affairs corresponding to p was brought about by a causal chain that makes no use of a's desires, beliefs, abilities, and the like.[9]

[8] Strictly speaking, there should be a temporal reference as well, for it is eminently plausible to think that a proposition might be unavoidable for an agent at one time, but not unavoidable for that same agent at some other time. To simplify the presentation, I have refrained from including such temporal indices in what follows. For the same reason, I have also adopted the common practice of letting some symbols stand indifferently for sentences or names—compare the use of H and L in (22) and (23).

[9] Slote implies only that something *like* (D1) is the appropriate analysis—he acknowledges that it "leaves out some factors relevant to unavoidability." See Slote, p. 19, fn 17. Such an analysis of unavoidability has been implicitly forwarded in recent years by several other compatibilists as well. See Alvin Goldman, *A Theory of Human Action* (Englewood Cliffs, N.J.: Prentice-Hall, 1970), pp. 84/5; Anthony Kenny, *Will, Freedom and Power* (New York: Barnes & Noble, 1975), p. 153; and Daniel Dennett, *Elbow Room* (Cambridge, Mass.: MIT Press, 1984), pp. 118, 127/8.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

If (D1) is employed, though, the Argument is obviously a failure, because (D1) renders rule of inference [B], which justifies the move from (1) and (2) to (3), patently invalid. For (given determinism), although Elfreth's current character traits clearly play no causal role in the explanation either of the laws of nature linking *S* to *A* or of *S* itself, they surely do play a role in bringing about Elfreth's act of iguana petting. As Slote puts it, "even if our deliberate actions are determined, that determinism operates, nonfatalistically, by means of (causal chains involving) the approximately coeval desires, abilities, character, and beliefs of human agents" (19). In short, if 'U' is analyzed via (D1), then, assuming determinism, (1) and (2) turn out clearly true, but (3) turns out clearly false. Hence, (D1) fails our first condition for an adequate analysis of unavoidability, for it renders the Argument invalid.

Slote sees the failure of (D1) as implying that the Argument itself is a failure, but such a conclusion is (to put it mildly) premature; for the failure of *one* purported analysis of unavoidability to satisfy our three conditions by no means implies that *no* analysis can succeed. Slote's focusing upon (D1) is all the more surprising given the fact that those who have publicly endorsed the Argument have not kept silent regarding the notion of unavoidability they think is involved. Peter van Inwagen, in fact, has offered us the means to fashion two definitions of the concept.[10] The first can be stated as follows:

(D2) $Uap =_{df}$ For any state of affairs x such that a can (i.e., has the power to) bring about x, it is necessarily the case that, if a brings about x, then p.

Suppose we allow 'Pax' to abbreviate "a can (has the power to) bring about x," and 'Cax' to abbreviate "a brings about x"; (D2) might then be symbolized:

(D2) $$Uap =_{df} \forall x[Pax \supset (Cax \Rightarrow p)]$$ [11]

[10] Van Inwagen talks not of *unavoidable* propositions, but of propositions that we can *render false*. But the two notions are clearly interdefinable, for to say that an agent *s* can render a proposition *p* false is just to say that *p* is *not* unavoidable for *s*; thus, for any analysis of rendering false, the negation of that analysans offers us a corresponding analysis of unavoidability. It might also be noted that each of these concepts is most plausibly seen as applying only to *true* propositions.

[11] See van Inwagen, "Reply to Narveson," *Philosophical Studies*, XXXII, 1 (July 1977): 89–98, p. 93. The actual definition of rendering false offered here by van Inwagen would yield

$$Uap =_{df} \forall x[Pax \supset \sim(Cax \Rightarrow \sim p)]$$

as our definition of unavoidability. On this definition, though, unavoidability would not seem to be demonstrably agglomerative, for the compatibilist will presumably say that, provided a is a free agent, we can produce a counterexample to (α) by

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

On the level of prima facie plausibility, (D2) would surely strike many as superior to (D1); for it does seem promising to think of something whose truth would be logically guaranteed by *anything* Elfreth might bring about as something that is indeed unavoidable for Elfreth. Furthermore, from the incompatibilist's point of view, (D2) possesses an even more significant advantage over (D1): whereas (D1) failed to make the Subargument valid, (D2) passes the test of validity. For, if 'U' is defined as (D2) suggests, then unavoidability is obviously both agglomerative (if p is entailed by anything I might bring about, and the same goes for q, then clearly the conjunction of p and q is entailed by anything I might bring about) and closed under entailment (for, if anything I might bring about entails p, and p entails q, then anything I might bring about entails q). Given (D2), then, both (α) and (β) hold, and we can be assured of the Subargument's validity.[12]

Unfortunately, (D2) fares less well when we turn to our second criterion for a successful definition—viz., its ability to render the Argument sound. In fact, given (D2), it seems evident that neither (1) nor (2) is true. For, if there is a state of affairs which Elfreth can bring about, surely there is a state of affairs which he can bring about such that his bringing it about *entails* neither that S implies A nor that the world was once as S says it was. So, although a definite improvement over (D1), (D2) is not the definition that the incompatibilist needs.

Perhaps the defect in (D2) can be corrected by merely building the past and the laws of nature into our definiens. A proposition that is unavoidable for Elfreth may not be, simply, the logical consequence of anything he brings about—that is the moral of (D2)'s failure—but will it not be the logical consequence of anything he brings about *given that* the world has had a certain *history* and that certain *laws of nature* hold? Reflections of this sort lead to a second analysis of unavoidability suggested by van Inwagen, namely:[13]

---

replacing p with H (the history of the world) and q with L (the laws of nature). Hence, I have changed the consequent of the definiens to $Cax \rightarrow p$, so as to produce the more formidable (D2).

[12] (D2) would also satisfy the relevance criterion, for it would be eminently unreasonable to accept (3) according to (D2) and also hold that Elfreth could have refrained from petting his iguana.

[13] See his *An Essay on Free Will*, p. 68. Van Inwagen's definition of rendering false is disanalogous to (D3) in two major respects: (i) it is couched in terms of an agent's "power to arrange or modify the concrete objects that constitute his environment," rather than in terms of one's power to bring about states of affairs; and (ii) it makes no explicit reference to the laws of nature in the definiens. Since the change in locution noted in (i) seems unmotivated and since talk of bringing about

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(D3) $Uap =_{df}$ For any state of affairs x such that a can bring about x, it is necessarily the case that, if a brings about x (given the laws of nature and the history of the world), then p.

Let 'H' stand for a proposition specifying the history of the world. (Intuitively, one might think of this as a conjunctive proposition, the conjuncts of which are, like S, descriptions of the world at particular times.)[14] Similarly, let 'L' stand for a proposition expressing all the laws of nature. (D3) might then be presented more perspicuously as

(D3) $$Uap =_{df} \forall x\{Pax \supset [(H \,\&\, L \,\&\, Cax) \Rightarrow p]\}$$

Is (D3) a genuine improvement upon (D2)? Let us consider how well it satisfies our three criteria of validity, soundness, and relevance.

First, would (D3) make the Subargument valid? That is, would (α) and (β) both be true given (D3)? Consider first (α). Suppose Uap and Uaq; would $Ua(p \,\&\, q)$ follow? Clearly, it would. For, given (D3), 'Uap' is identical with

(4) $$\forall x\{Pax \supset [(H \,\&\, L \,\&\, Cax) \Rightarrow p]\}$$

whereas 'Uaq' is identical with

(5) $$\forall x\{Pax \supset [(H \,\&\, L \,\&\, Cax \Rightarrow q]\}$$

Now, from (4) and (5), one can easily derive

(6) $$\forall x\{Pax \supset [(H \,\&\, L \,\&\, Cax) \Rightarrow (p \,\&\, q)]\}$$

which, given (D3), is identical with '$Ua(p \,\&\, q)$'. So (α) is indeed true—'U' is indeed agglomerative—if (D3) is our definition.

What, then, of (β)? Does it also turn out to be true given (D3)? Yes, it does. Suppose Uap and $p \Rightarrow q$. As we have seen, 'Uap', given (D3), is identical with (4). Instantiating, we can derive

(7) $$Paz \supset [(H \,\&\, L \,\&\, Caz) \Rightarrow p]$$

---

states of affairs seems clear enough, I have not followed van Inwagen's new way of speaking; however, readers who wish to may do so, I think, without substantially affecting the discussion of (D3). As for (ii), a definiens which *does not* refer to the laws of nature would be open to objections very similar to the Nostradamus counterexample mentioned by van Inwagen (67). Finally, in order to assure agglomerativity, I have made a change in the definiens of the same type as that mentioned above in note 11.

[14] I am presuming that the history of the world is inclusive enough so that compatibilists and incompatibilists may differ over whether that history plus the laws of nature determine the future, but is *not* so inclusive as to make that history *by itself* determinative of the future. Hence, the history of the world is not to be thought of as including future-tense propositions. For a detailed explication of the Ockhamist notion of history being presupposed here, see Alfred J. Freddoso, "Accidental Necessity and Logical Determinism," this JOURNAL, LXXX (May 1983): 257–278.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

from (4). But (7), combined with our supposition that $p \Rightarrow q$, patently entails

(8) $$\mathrm{P}az \supset [(H \,\&\, L \,\&\, \mathrm{C}az) \Rightarrow q]$$

which (via universal generalization) yields

(9) $$\forall x\{\mathrm{P}ax \supset [(H \,\&\, L \,\&\, \mathrm{C}ax) \Rightarrow q]\}$$

But, given (D3), (9) is identical with '$\mathrm{U}aq$'. So (D3) assures us that, if $\mathrm{U}ap$ and $p \Rightarrow q$, then $\mathrm{U}aq$—i.e., it assures us that 'U' is closed under entailment.

Since (α) and (β) both turn out true on (D3), there can be no doubt that (D3) guarantees the Subargument's validity. Does it, though, also guarantee the Subargument's soundness? That is, does it render (1) and (2) beyond reasonable doubt if determinism is assumed? Consider (1) first. For (1), interpreted via (D3), to be false, there would have to be some state of affairs (call it z) which Elfreth can bring about such that

(10) $$\sim[(H \,\&\, L \,\&\, \mathrm{C}ez) \Rightarrow (S \supset A)]$$

Now, if determinism is true, then the history of the world and the laws of nature together entail that Elfreth pet his iguana at t. So the determinist will affirm

(11) $$(H \,\&\, L) \Rightarrow A$$

But, of course, since

(12) $$A \Rightarrow (S \supset A)$$

it follows that the determinist is also committed to

(13) $$(H \,\&\, L) \Rightarrow (S \supset A)$$

Since antecedents of entailments can always legitimately be strengthened, (13) in turn entails

(14) $$(H \,\&\, L \,\&\, \mathrm{C}ez) \Rightarrow (S \supset A)$$

But (14) is simply the negation of (10). Since, given (D3), the affirmation of (10) is required if one is to deny (1), and, since, as we have just seen, anyone assuming determinism to be true is committed to affirming (14) and thus to rejecting (10), it follows that, given determinism, (1) is indeed true if (D3) is our analysis of unavoidability.

Is (2) similarly beyond dispute? Clearly, it is. For, given (D3), (2) could be false only if there were some state of affairs z which Elfreth can bring about such that

(15) $$\sim[(H \,\&\, L \,\&\, \mathrm{C}ez) \Rightarrow S]$$

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

But, since *H* specifies the complete history of the world, it will surely include that description of the world at a certain time which *S* provides. In other words, *H* by itself will entail *S*. But then, of course, *H* conjoined with anything whatsoever will also entail *S*. So there is simply no way in which (15) could be true. And, since (2) could be false only if (15) were true, it follows that (D3) does indeed guarantee the truth of (2).

Thus far, (D3) has been most accommodating to the incompatibilist. As we have seen, provided that we understand the Argument as speaking of unavoidability in the sense offered by (D3), we can demonstrate conclusively that both of the inferential rules presupposed by the Subargument are valid, and that both of the premises it employs are true if determinism is true. But, if the line of reasoning given in the Subargument is thus valid and sound, the compatibilist has little choice but to concede that (3) is true as well. That is, the compatibilist is committed to saying that, in the sense given by (D3), it is indeed unavoidable for Elfreth that he pet his iguana at *t* if determinism is true.

Rash incompatibilists might be tempted to claim victory at this point, for (they might claim) if Elfreth's act of iguana petting was unavoidable, he could not have refrained from performing it, and, if he could not have refrained, his action was not free. Circumspect compatibilists, however, would not surrender so easily. True, they might say, (D3) does indeed satisfy the validity and soundness criteria; but does it satisfy the relevance criterion? All that has been proved thus far is that it was unavoidable *in the sense given by (D3)* that Elfreth pet his iguana. But how do we know that unavoidability *in this sense* is inimical to freedom? Might it not be that, even though (3) is true *in the sense given by (D3)*, it is also true that Elfreth could have refrained from petting his iguana? The use of the word 'unavoidable' might *incline* us to think that the truth of (3) is injurious to Elfreth's freedom, but that merely shows the rhetorical craftiness of the incompatibilist. Is there any *reason* to regard (3) interpreted according to (D3) as relevant to the question of freedom?

The incompatibilist might try to show that it is. What the compatibilist is requesting is a demonstration that it follows from (3) that Elfreth could not have refrained from petting his iguana. Now, presumably, Elfreth could have refrained from petting his iguana if and only if Elfreth had the power to bring about that state of affairs which corresponds to $\sim A$ (i.e., the state of affairs of its not being the case that Elfreth pets his iguana at *t*). Let '$\bar{A}$' stand for this state of affairs. What the compatibilist wants, then, is a reason to think that

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(16) $$\sim Pe\bar{A}$$

does indeed follow from (3). Given (D3), (3) is identical with

(17) $$\forall x\{Pex \supset [(H \,\&\, L \,\&\, Cex) \Rightarrow A]\}$$

But why think that (17) entails (16)? Well, (17) obviously implies

(18) $$Pe\bar{A} \supset [(H \,\&\, L \,\&\, Ce\bar{A}) \Rightarrow A]$$

If we could show that the consequent of (18) is false, it would follow that its antecedent is also false; and, since its antecedent it simply the negation of (16), that would suffice to show that (16) does follow from (17).

So the challenge for the incompatibilist is to show that no one could reasonably accept the consequent of (18), namely

(19) $$(H \,\&\, L \,\&\, Ce\bar{A}) \Rightarrow A$$

But why not? Well, a libertarian might argue, if Elfreth brings about $\bar{A}$, it clearly would have to be the case that $\sim A$. In other words,

(20) $$Ce\bar{A} \Rightarrow \sim A$$

is obviously true. But (20) clearly implies

(21) $$(H \,\&\, L \,\&\, Ce\bar{A}) \Rightarrow \sim A$$

Given the truth of (21), (19) could be true only if these two conditionals' common conjunctive antecedent were impossible—i.e., only if

(22) $$\sim\Diamond(H \,\&\, L \,\&\, Ce\bar{A})$$

Since (22) is false, it follows that (19) is false as well.

Unfortunately, the last step in this argument is little more than blatant libertarian bluster. For why should compatibilists concede that (22) is false? After all, (22) is equivalent to

(23) $$(H \,\&\, L) \Rightarrow \sim Ce\bar{A}$$

But, far from thinking that (23) is false, will not compatibilists see (23) as a simple and obvious consequence of the deterministic thesis they espouse (or at least countenance espousing)? If, as a matter of fact, Elfreth *did* pet his iguana at t, then determinism tells us that the history of the world conjoined with the laws of nature entails that he pet his iguana—that is, it tells us that

(24) $$(H \,\&\, L) \Rightarrow A$$

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is true. Since it is clearly the case that A is incompatible with Elfreth's bringing about $\sim A$, the compatibilist will also insist that

(25) $$A \Rightarrow \sim \mathrm{C}e\bar{A}$$

But (24) and (25), of course, patently imply (23) and its equivalent, (22). So compatibilists would seemingly be well within their rights to deny that (22) is false and, hence, to deny that they have been given any reason to think that (17) implies (16). And, in the absence of such a reason, they would no doubt remind us, the truth of (3) would appear to be fully compatible with the truth of compatibilism.

Is (D3), then, ultimately a failure? The incompatibilist might try one more strategy to show that (D3) passes the relevance criterion —i.e., that (17) does entail (16). For consider the following principle:

(26) $$\forall a \forall p[\mathrm{P}ap \supset \Diamond(H \,\&\, L \,\&\, \mathrm{C}ap)]$$

According to this principle, a necessary condition of an agent's having the *power* to bring about a certain state of affairs is that it be *possible,* given the history of the world and the laws of nature, that the agent bring about that state of affairs. Now, such a principle, the incompatibilist might say, is clearly true; for what sense would it make to speak of an agent's having the *power* to do something if the agent *could not actually* do it? In the case of Elfreth and his iguana petting, (26) tells us that

(27) $$\mathrm{P}e\bar{A} \supset \Diamond(H \,\&\, L \,\&\, \mathrm{C}e\bar{A})$$

Now, the compatibilist was able to accept (17) while rejecting (16) only by insisting that (22) was true. But (22) is simply the negation of the consequent of (27). By endorsing (22), then, compatibilists have committed themselves to denying the antecedent of (27). However, to deny the antecedent of (27) just *is* to affirm (16). So, the incompatibilist might conclude, there is no way of escaping the conclusion that (17) entails (16); and, if it does, then (D3) passes the relevance test and renders the Argument a success.

This argument shows conclusively that, if (26) and (27) are acceptable principles, then (D3) satisfies each of our three conditions; if they are *not* acceptable, though, it is difficult to see how the relevance test can be met by (D3). Now, it might be thought that this reliance upon (26) and (27) is in no sense an essential feature of the Argument; further investigation, one might suspect, would provide us with a definition and an Argument that presupposed the truth of neither (26) nor (27). In fact, however, the more we canvass the plausible candidates for a definition of 'U', the more quixotic it

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

appears to think that the Argument can get by without something like (26) and (27) as presuppositions.

We might, for example, try exchanging talk of *entailment* in our definiens for talk of *counterfactual* or *subjunctive* implication, thereby arriving at definitions such as

(D4) $Uap =_{df}$ for any state of affairs x such that a can bring about x, if a were to bring about x, then it would be the case that p.

or

(D5) $Uap =_{df}$ for any state of affairs x such that a can bring about x, if a were to bring about x (given the laws of nature and the history of the world), it would be the case that p.

Given (D4), however, (2) is no longer uncontroversially true; thus, (D4) fails the soundness test.[15] And (D5), though yielding an argument that is demonstrably valid and sound, stands in the same position vis-à-vis the relevance test as (D3): it seems to pass only if (26) and (27) can safely be assumed.

Other definitions of 'U' might be ventured, but none that I have seen or devised manages to give us a successful Argument without employing something along the lines of (26) and (27).[16] Can we *prove* that the Argument tacitly assumes such theses? I doubt it. But, until we are presented with a version of the Argument the success of which *does not* depend upon the acceptability of (26) and (27), it seems fair to assume that they are indeed required.

[15] Most compatibilists, I think, would reject (2) interpreted according to (D4) because they would hold that Elfreth did have the power to refrain from petting his iguana, but would actually have refrained only if H had been different. On the other hand, compatibilists who feel that worlds with different laws are closer to our world than are worlds with different histories would presumably be inclined, given (D4), to reject (1) rather than (2). It should be noted that neither of these responses need be based on anything so questionable as a wholesale rejection of "backtracking" counterfactuals. For a compatibilist defense of such counterfactuals, see Jonathan Bennett, "Counterfactuals and Temporal Direction," *Philosophical Review*, XCIII (January 1984): 57–91.

[16] To mention one prominent example: Wiggins (45/6) suggests a definition such as

(D6) $Uap =_{df}$ p holds in every possible world whose history and laws of nature are indistinguishable from the history and laws of nature of the actual world.

Although such a definition evidently yields a valid and sound Subargument, it would appear to pass the relevance test only if a very close relative of (26), namely,

(26*) $\forall a \forall p[Pap \supset \Diamond(H \,\&\, L \,\&\, p \text{ occurs})]$

can be taken for granted. And, of course, compatibilist reservations regarding (26) are equally applicable to (26*).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

III

Our evaluation of the Argument appears, therefore, to depend crucially upon our verdict regarding (26) and (27). Now, it might be felt that, if these two theses were established, the triumph of the Argument would be unquestioned; and, in a sense, this feeling is surely on target. In another sense, however, it is not. For, if something like (27) can safely be assumed, we can easily and directly refute compatibilism, and refute it without engaging in any complex prestidigitations involving analyses of unavoidability. In fact, we can refute compatibilism without making any reference to unavoidability at all.

Suppose, as the compatibilist thinks we consistently can, that determinism is true and that Elfreth freely pets his iguana at t. If so, then Elfreth's iguana petting was the metaphysically necessary consequence of the history of the world prior to t and the laws of nature. So, letting 'D' abbreviate the thesis that determinism is true, we know that

(28) $$(\mathrm{D} \,\&\, \mathrm{A}) \supset [(H \,\&\, L) \Rightarrow A]$$

Since, as we have already noted,

(25) $$A \Rightarrow \sim \mathrm{C}e\bar{A}$$

is obviously true, it follows from (28) that

(29) $$(\mathrm{D} \,\&\, A) \supset [(H \,\&\, L) \Rightarrow \sim \mathrm{C}e\bar{A}]$$

which is equivalent to

(30) $$(\mathrm{D} \,\&\, A) \supset \sim\Diamond(H \,\&\, L \,\&\, \mathrm{C}e\bar{A})$$

Now, suppose it were beyond dispute that

(27) $$\mathrm{P}e\bar{A} \supset \Diamond(H \,\&\, L \,\&\, \mathrm{C}e\bar{A})$$

From (27) and (30) we could then infer

(31) $$\mathrm{P}e\bar{A} \supset \sim(\mathrm{D} \,\&\, A)$$

Since compatibilists want to affirm Elfreth's power to do other than he did, they will affirm the antecedent of (31), and so they have no choice but to affirm its consequent—i.e.,

(32) $$\sim(\mathrm{D} \,\&\, A)$$

And, since A is beyond dispute, compatibilists would be left with no choice but to reject D—i.e., to renounce determinism.

Even if (27) were an acceptable starting point for discussion, then, the Argument would have to be judged to be little more than a

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

convoluted superfluity. True, it would succeed in refuting compatibilism, but it would do so at the cost of introducing a concept (unavoidability), along with definitions and rules of inference for that concept, which were totally unnecessary to effect that refutation. Such an Argument could hardly be seen as throwing light on the issue, or as genuinely advancing the discussion.

Of course, if (26) and (27) were *not* beyond dispute—and, a fortiori, if they were themselves instances of the very type of thesis over which compatibilists and incompatibilists have traditionally wrangled—the Argument would have even less right to be seen as a step in the right direction. Unfortunately for the incompatibilist, this seems to be the case. For, traditionally, compatibilists have tried to reinforce their position by offering conditional analyses of power—i.e., analyses according to which someone like Elfreth had the power to refrain from petting his iguana just in case he would have refrained had such and such been the case. In filling out such an analysis, compatibilists have not felt constrained to guarantee that the consequent of (27) be true—i.e., that Elfreth's refraining might actually have occurred even if nothing else had been different. On the contrary, J. S. Mill, Moritz Schlick, and others have made a point of saying that one who *could* have done otherwise *would* have done otherwise only if some element in the situation (the agent's beliefs, desires, and the like being those most often cited) *had* been different. Those who favor such conditional analyses, then, would not view either (27) or the more general (26) as obvious truths; they would see them, rather, as evident falsehoods.

Now, it is fair to say that no particular way of completing this conditional analysis has garnered anything approaching universal acceptance among compatibilists. It is also fair to say that the most obvious ways of trying to complete it in a way amenable to compatibilists have been subjected to rather severe criticism by J. L. Austin, C. D. Broad, Roderick Chisholm, and others.[17] But it would surely constitute the height of unfairness to say that we can therefore take it as an indisputable starting point in our discussions of freedom that any such analysis must fail; and that is just what we would be doing if

[17] See Austin, "Ifs and Cans," in Bernard Berofsky, ed., *Free Will and Determinism* (New York: Harper & Row, 1966), pp. 295–322; Chisholm, "J. L. Austin's Philosophical Papers," *ibid.*, pp. 339–345; and Broad, "Determinism, Indeterminism and Libertarianism," *ibid.*, pp. 135–159. It is worth noting that not all compatibilists look favorably upon conditional analyses; see, in particular, Keith Lehrer, "An Empirical Disproof of Determinism?", pp. 175–202 in his anthology cited above. Still, there can be little doubt that even Lehrerian compatibilists would hardly find (26) and (27) acceptable.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

we were to presuppose that (26) and (27) are beyond dispute. (26) and (27) may well be true—indeed, I believe they are true—but to expect the compatibilist to be moved by an argument that relies upon them is surely to expect a bit too much.

This is not to imply that (26) and (27) have nothing to be said for them or are of no dialectical value. The incompatibilist might well point to them and ask whether, on reflection, we do not just see that (26) *does* express a necessary condition of an agent's having the power to perform a certain action. Some compatibilists—or, more likely, some of those as yet uncommitted to a position—might be led in this manner to accept (26) and (27) and, consequently, to embrace incompatibilism. But it seems doubtful that many compatibilists would be moved in this way; and, if they are not, it is hard to see how the incompatibilist can claim that their position is refuted. If there were some non-question-begging way to prove (26) or if it were a proposition that was manifestly self-evident, things might be different. But (26) does not appear to be self-evident (many people seem to understand it well enough and still reject it), and it is difficult to imagine how one might go about demonstrating it to one who understood it, yet honestly doubted it. Thus, an argument against compatibilism which depends upon (26) cannot be judged to be a refutation of that position. And so the Argument from Unavoidability would appear to be a failure.

IV

Can we conclude, then, that compatibilism is irrefutable? Of course not. For all that has been said here, there may well be some other argument that, relying not on (26) or (27) but only on premises that any reasonable compatibilist would accept, succeeds in showing determinism to be incompatible with human freedom. Commendable as the search for such an argument surely is, it may well be that the incompatibilist's best means of fostering philosophical progress lies in a different direction—that of developing a more finely articulated version of incompatibilism and responding to objections that compatibilists or others raise against it.[18] This more positive approach may lack much of the glamor and excitement that attend the search for a refutation of compatibilism, but I suspect that it will in the end prove more fruitful.

THOMAS P. FLINT
University of Notre Dame

[18] Much of this work has already been done, most notably by Richard Taylor and Chisholm; see especially Taylor's *Action and Purpose* (Englewood Cliffs, N.J.: Prentice-Hall, 1966), and Chisholm's *Person and Object* (LaSalle, Ill.: Open Court, 1976).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

DECIDING TO BELIEVE WITHOUT SELF-DECEPTION*

DAVID HUME noticed (as have many others) that believing is not directly subject to our conscious control. His account suggests that it is a matter of contingent fact that one's beliefs do not respond to one's will, similar to the way in which hiccoughs do not respond to one's will. Bernard Williams[1] has persuasively argued that our inability to believe at will is not a contingent fact, but a consequence of the conceptual relations among the notions of belief, evidence, and truth. To believe *p* is to believe that *p* is true. But, if I could simply will myself to believe *p*, then I could do so whether *p* is true or not, and I would know that I could do so. Presumably the notion of trying to will myself to believe something would only occur to me when I find the evidence for *p* insufficient to warrant belief in the normal way. So, if it were possible for me knowingly to believe at will, it would have to be possible for me to believe that *p* is true, knowing all the while that the belief is unwarranted by the evidence, that the belief was acquired at will, and that beliefs can be so acquired irrespective of their truth or falsity. Williams rightly judges that this borders on incoherence, and concludes that "With regard to no belief could I know—or, if all this be done in full consciousness, even suspect—that I had acquired it at will" (148).

Having rightly ruled out the possibility of believing "at will, just like that," Williams grants that "there is room for application of decision to belief by more roundabout routes" (149). But the pursuit of belief by such "roundabout routes" is, he says, deeply irrational and can succeed only by means of self-deception. Presumably one would at least have to deceive oneself about the origins of one's belief—i.e., about the fact that one lacked warranting evidence for one's belief and had arrived at one's belief as a consequence of having voluntarily undertaken a program of belief inducement via roundabout routes. This is how Jon Elster[2] interprets Williams's position in his recent discussion of the issue. Elster approvingly states the principle as follows: "one cannot both believe *p* and believe that one's belief that *p* stems from a decision to believe" (50).

David Pears[3] takes a similar position in his discussion of a hard case

* I am grateful to Michael Phillips, Amelie O. Rorty, Peter Skagestad, Lilly-Marlene Russow, and Michael McPherson for helpful comments at various stages of this paper's development.

[1] "Deciding to Believe," in *Problems of the Self* (New York: Cambridge, 1973).

[2] *Ulysses and the Sirens* (New York: Cambridge, 1979).

[3] "Freud, Sartre and Self-Deception," in Richard Wollheim, ed., *Freud: A Collection of Critical Essays* (Garden City, N.Y.: Anchor Books, 1974), pp. 97–112.

0022-362X/87/8408/0441$00.60 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

in which an individual believes not-*p* but finds that belief unacceptable and sets out to induce in himself a belief that *p*. After recounting strategies by which one might try to manipulate oneself into holding the desired belief, he states, "At some point in your progress it must become impossible for you to see through what you are doing to your motive for doing it" (108). Later he reiterates, "if this plan is going to succeed, there must come a stage in its execution at which you cannot review it in a way that brings in your motive" (*ibid.*).[4]

Each of these philosophers holds that an effort to induce in oneself a belief in the truth of some proposition that one believes to be false (or, an easier case, a proposition for which one lacks warranting evidence) can succeed only if one manages, somewhere along the way, to forget that one is engaged in such an effort. Although this view has strong intuitive appeal, I think that it is false, and that it can be shown false by example. I suspect that Williams's, Elster's, and Pears's acceptance of the view stems, first, from an underestimation of the extent to which nonevidential factors can influence belief, and, secondly, from a disregard of the fact that one's criteria for evidential warrant are themselves beliefs amenable to manipulation in "roundabout" ways. As the following example shows, it is indeed possible to believe *p* while believing that one's belief stems from a motivated decision to believe *p*, so long as one manages manipulatively to alter not only one's beliefs about the truth or falsity of *p* but also one's criteria for evidential warrant with regard to such beliefs.[5]

> Nick is a bright, very ambitious biology major at a small fundamentalist college. He aspires to professional success as a biologist, and is willing to go to great lengths to attain it. He rightly suspects that his professional prospects are minimal so long as he maintains and avows his hitherto lifelong creationist beliefs. He has considered remaining a closet creationist while publicly avowing evolutionary theory, but a life of hypocrisy and deceit is an unappealing prospect. Overall, he judges that the best solution to his problem would be for him to believe that evolution is the correct theory of the development of life forms on earth. So he decides to believe it.

[4] In a passage *a propos* of the case to be considered, Pears qualifies his claim slightly by saying: "In the later stages either my original motives must be screened, or it will come through in a version which does not include my rational tendency to believe not-*p*" (107). Pears does not explain this, but my example may be a case of what he has in mind. (See footnote 6.)

[5] This example was suggested to me by Peter Skagestad in an unpublished paper on Nietzsche's *Uses and Abuses of History*. Overall, the ideas in this paper have developed in stimulating discussions with Skagestad.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Nick believes (and always has) that revealed scripture, taken literally, is the ultimate arbiter of disputed questions, so it will be difficult for him to induce in himself the desired belief. But he knows something about the way in which nonevidential factors influence beliefs and he works out a project of belief inducement based upon that knowledge. He knows enough about psychology to know about dissonance-reduction mechanisms of the mind which tend to bring belief into consonance with behavior. He knows enough social psychology to know the extent to which our beliefs are influenced, in nonevidential ways, by the people whom we admire and with whom we spend time. He knows enough hermeneutics and philosophy of science to know that the ways in which evidence is interpreted—and even what counts as evidence—is in great part a function of one's guiding theory and interpretive hypotheses. And, finally, he knows enough to know that our selective attention to evidence and employment of interpretive hypotheses are largely a matter of habits—habits which, like other habits, can be cultivated over time.

So, in a move designed both to further his career directly and to bring it about that he accepts evolutionary theory, he decides to go to Harvard for graduate training. Though he believes evolutionary theory to be false, he has decided, on prudential grounds, to believe that it is true, and he is confident that the prestige of his instructors, various social pressures, and the internal dissonance brought on by the need constantly to avow and utilize evolutionary assumptions (in class, in exams, etc.) will ultimately induce in him the desired belief.

According to Williams's and Elster's thesis, Nick will be unable to succeed in his project of belief inducement unless he manages to forget or to deceive himself about his motives and about the origins of his belief. On this view, so long as he remembers that his belief is a consequence of his having employed a roundabout strategy of belief inducement, relying on non-evidential mechanisms and influences, he will be unable really to believe in the truth of the theory of evolution. But to see the falsity of this thesis, we need only fill in the rest of the story:

Six years later, Nick emerges from graduate school, Doctorate in hand, fully versed in and accepting of contemporary biological theory. He remembers quite well the entire story of how he had decided to come to Harvard in hopes that he would be seduced into believing that which he then thought to be false. He smiles inwardly at the benighted bumpkin he was—so benighted that he had thought evidential considerations insufficient to warrant belief in the truth of something as *obviously* true as evolutionary theory.

The point of the story is that Nick has successfully carried out his decision to believe p and is fully aware that his belief is a consequence

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of a strategically planned "roundabout" program of belief inducement. He has not forgotten anything, and has no need to deceive himself about anything. Thus, Williams's and Elster's thesis is shown to be false.

One might argue that Nick has indeed forgotten something—he has forgotten that revealed scripture is the final arbiter of disputed questions. But, in fact, he has not forgotten this; he holds simply that it is false and that he was mistaken when he held it to be true. Perhaps, then, he has forgotten the evidence that once led him to hold his prior beliefs. But, in fact, he has not forgotten it at all; it is just that now he is inclined to think that evidence very weak and to think that his prior beliefs were mostly without evidential support, resting almost entirely upon ignorance, social pressure, superstition, and unjustified credence given to the views of his fundamentalist parents and teachers. In short, he has not forgotten anything about the past, though his present account of his prior epistemic state is different from the account that he would have offered previously. The story that he now tells about his previous epistemic position is a redescription of that position, but not one which requires either selective forgetfulness or self-deception.

It is important to note that there is a kind of symmetry between Nick's earlier perspective upon his later position and his later perspective upon his earlier position. Referring to him prior to implementation of his strategy of belief inducement as Nick_1, and upon graduation from Harvard as Nick_2, this symmetry can be described as follows: Nick_1 views his own beliefs as based upon warranting evidence and foresees Nick_2's beliefs to be the results of nonevidential causal influences (professors' prestige, social pressures), prejudicially selective habits of attention, and employment of tendentious interpretive hypotheses. Likewise, Nick_2 views his own beliefs as evidentially warranted and those of Nick_1 as the result of ignorance, distorted perspective, ritual belief reinforcement and other nonrational causal influences. The fact that each can "explain away" the beliefs and the evidential standards of the other provides sufficient "epistemic insulation" between the two to permit Nick_2 sincerely to believe *p* without having to forget or deceive himself about the fact that his belief is a consequence of Nick_1's decision to believe *p*. Perhaps this is what Pears had in mind when he suggested that "my original motive must be screened, or . . . come through in a version that does not include my rational tendency to believe not-*p*" (107).[6]

[6] See above footnote 4.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

In Nick's case, his original motive "comes through in a version" that includes a memory of having had a tendency to believe not-*p* and having thought it rational, but his present view of that tendency holds it not to have been rational. (Or, if perhaps it was rational given his evidence at the time, that evidence was woefully inadequate—so inadequate that it led him to distorted notions of rational warrant.)

In a sense, the suggested counterexample to Williams's thesis is but an application of the Duhem-Quine thesis, according to which one can hold *any* proposition true if one is willing (and able!) to make sufficient changes elsewhere in one's system of beliefs. Williams might hold that this sort of case is devoid of practical interest precisely because the individual, having had to change his standards of evidential warrant, has also had to change innumerable other beliefs —and this, it might be thought, would be so disruptive that no one would choose to do it. In response, it should first be said that this objection does not affect the conceptual import of the case. Moreover, the extent to which such strategic belief inducement will require alteration of other beliefs will vary widely from case to case. In some cases, massive overhauling of one's belief system might be required; in other cases, the effects of the change might be predictably slight or easily isolable.[7]

I would suggest, in conclusion, that rather than think of what Nick has done as a case of self-deception, we would do better to construe it along the lines of conscious character planning on an Aristotelian model. Suppose that I do not enjoy some *x*, but think that I would be a better person if I did. So I set about to engage in certain patterns of action, the predictable result of which will be my coming to enjoy *x*. In time, a *hexis* develops and I find myself enjoying *x*. In M. F. Burnyeat's[8] felicitous phraseology, I learn to enjoy *x* where that is "not sharply distinct from my learning that *x* is enjoyable." A cynic might say that I had deceived myself about the enjoyability of *x*, but a more charitable account would hold that I had successfully carried out a project of conscious character planning.

[7] It is, after all, quite possible to maintain different criteria of evidential warrant with regard to different aspects of one's life or different subject matters. All that is needed is either failure to notice that one is employing different criteria or some (perhaps quite vague) cover story about the appropriateness of the disparate criteria. Consider also recent suggestions that loyalty consists in part of a willingness to apply more stringent criteria of evidential warrant with regard to unflattering or maligning claims about those to whom one is loyal.

[8] "Aristotle on Learning to be Good," in A. O. Rorty, ed., *Essays on Aristotle's Ethics* (Berkeley, Calif.: University Press, 1980), p. 76.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Nick's motives are perhaps not so noble, but the structure of the case is comparable. He set out to engage in a pattern of action, the predictable result of which was his coming to believe a certain theory in biology. In time, a *hexis* developed, and he found himself believing the theory. He learned to believe the theory, where that is not sharply distinct from learning that the theory is believable. One might judge that Nick has deceived himself, but a more charitable account would hold that he has successfully carried out a program of belief acquisition via a roundabout route.[9]

Most of us would probably agree that Nick moved from an unwarranted belief in a false view to a warranted belief in a true one. But our judgment of the force of the example should not be thought to depend on this fact. If Nick had started out as an evolutionary biologist and had had sufficiently strong prudential motives for wanting to believe some fundamentalist/creationist doctrine, he could have done so, and could have done so without self-deception. One's sense for what is believable (otherwise known as one's criterion for epistemic warrant) can be altered in any number of nonrational ways. To the extent that one manages to bring about such an alteration, one's beliefs can be consciously manipulated without the need for self-deception.

J. THOMAS COOK

Rollins College

[9] As an aside, I might mention that those with Parfitean inclinations could interpret the apparent absence of self-deception in this case in a different way. It might be thought that there is no single Nick who is both deceiver and deceived, since $Nick_1$'s machinations not only produced certain beliefs, but brought into being a different (later) self to hold those beliefs. Thus, though $Nick_1$ did willfully bring it about that $Nick_2$ hold beliefs that $Nick_1$ held to be false, the deception involved here is not self-deception, for the deceiver and the deceived are quite distinct selves.

I have reservations about this way of talking, but the epistemic insulation between $Nick_1$ and $Nick_2$, coupled with the rather patronizing story that each has to tell about the other's epistemic status, suggests that they are ripe candidates for being considered different selves.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

Logic as Grammar. NORBERT HORNSTEIN. Cambridge, Mass.: MIT Press, 1984. xiv, 176 p. Cloth $22.50, paper $6.95.

Hornstein is concerned with properties of natural-language quantifiers and pronouns. The major contrast drawn is between quantifiers represented by 'a certain' and those represented by 'every'. According to Hornstein, the former can, whereas the latter cannot, coindex pronouns that occur (1) to their left, (2) outside their own clauses, or (3) across sentences in a discourse:

(1a) That *he* might be killed didn't bother *a certain soldier.*
(b) *That *he* might be killed didn't bother *every soldier.*

(2a) If Mary brings *a certain man* to the meeting, then *he* will make things difficult for the lecturer.
(b) *If John likes *every dog,* then *it* likes him.

(3a) Several men chased *a certain woman.* John caught *her.*
(b) **Every citizen* can own a gun. However *he* must register it.

Two further contrasts: According to Hornstein, 'a certain woman' unambiguously takes wide scope in (4a); whereas 'every man' can take either wide or narrow interpretive scope in (4b):

(4a) All men love a certain woman.
(b) Someone loves every man.

Finally, 'a certain'-quantifiers can take wide interpretive scope over matrix verbs when they occur as subjects of tensed subordinate clauses; 'every'-quantifiers, according to Hornstein, cannot:

(5a) Everyone believes that a certain woman is flighty.
(b) Someone believes that every woman is flighty.

In chapters 2 and 3 Hornstein claims that an element exhibiting any of the (positive) properties illustrated by 'a certain' will exhibit all five of those properties, and an element exhibiting any of the (negative) properties illustrated by 'every' will exhibit each of those properties. The former, type 1 quantifiers, are said to include definite descriptions and 'any' (uniformly construed as a universal quantifier). The latter, type 2 quantifiers, are said to include 'everyone', 'someone', ⌜every G⌝, and ⌜a G⌝.

* This review was written while I was serving as a visiting professor at the University of Washington.

0022-362X/87/8408/0447$00.90 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Hornstein uses this tidy classification to motivate a strongly nativist conception of semantic competence. According to this conception, understanding a sentence involves using internalized linguistic rules to derive a psycho-logical form[1] from its internal syntactic representation. Innate constraints on rules and the resulting representations are thought to provide the learner with a highly restricted system for analyzing linguistic data. Exposure to data like (3)–(4) fixes the psycho-logical representation of different quantifiers, which in turn is supposed to determine their other interpretive properties. Only in this way, Hornstein believes, can one explain how speakers acquire such a rich system of semantic competence from the apparently impoverished data available in (first) language learning.

Hornstein's analysis is a version of Noam Chomsky's recent Government and Binding approach, with some extensions and modifications. Hornstein regards it not just as a psychological hypothesis, but also as a theory of meaning. This is striking since there is no room in it for an account of truth conditions or for any other mapping of linguistic expressions onto the extralinguistic reality they represent. Hornstein thinks this is a virtue. He argues that familiar semantic approaches do not explain what semantic competence is and how it is acquired, and concludes that these approaches have no importance for theories of meaning. In their place we get a mapping from English to (uninterpreted) mentalese. Hence the title of the book (with "logic" signifying a central aspect of meaning).

The book contains significant problems of detail. More important, three fundamental difficulties stand out: the quantifier taxonomy is inaccurate; vagueness and equivocation eviscerate the theory's predictive content; and the rejection of truth conditions violates the theory's own requirements.

First, some obvious counterexamples to the taxonomy:

(6a) That he might be laughed at bothered *at least one performer.*
(b) That he might be laughed at bothered *a performer I represent.*
(c) The fact that they are trailing in the polls bothers *several candidates.*[2]

Like 'a certain', each of these quantifiers can coindex a pronoun to its left. Thus, Hornstein is committed to the claim that each is a type 1 quantifier that behaves like 'a certain' regarding all the properties

[1] The representation is supposed to be a logical form that is psychologically real—hence, my terminology, 'psycho-logical form'.

[2] The relevant interpretation of (6c) requires several candidates to satisfy "the fact that x is trailing in the polls bothers x."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

illustrated by (1)–(5). These quantifiers, however, are not unambiguously interpreted as having wide scope:

(7a) Every director wants to sign at least one performer.
(b) Every director wants to sign a performer I represent.
(c) Every columnist hates several candidates.

Similar results can be obtained with other properties and quantifiers.

Even more troublesome is the pervasive indefiniteness at the heart of the theory. The interpretive differences between type 1 and type 2 quantifiers are supposed to result (primarily) from their positions in psycho-logical form. Type 2 quantifiers are sentential operators in psycho-logical syntax; type 1 quantifiers, like names, are terms. The key interpretive notion characterizing such terms is "interpretive independence." We are told that interpretively independent elements "are relatively insensitive to their 'logical' environments" (44), that they "are interpreted in a manner functionally independent of the interpretation of other logical operators in the clause" (30), and that in general this means that they are "interpreted as if they had widest possible scope" (44). No more precise characterization is given.

One place where this leads to trouble is in the analysis of 'any' as an interpretively independent universal quantifier. Although many sentences with 'any' can be represented by treating it as a wide-scope universal quantifier, some, like (8), cannot:

(8) Someone doesn't like anyone.

Hornstein handles this case by saying that interpretively independent quantifiers are interpreted as if they were in branching structures. Thus, (8) has the interpretation (9a), which is equivalent to (9b).[3]

(9a) $\left.\begin{array}{l}\text{someone } x\\ \text{everyone } y\end{array}\right\rangle$ [not (x likes y)]

(b) someone x, everyone y [not (x likes y)]

[3] The analysis of (i) is given by the branching structure (ii), which is equivalent to (iii):

(i) Everyone dislikes a certain person.

(ii) $\left.\begin{array}{l}\text{someone } x\\ \text{everyone } y\end{array}\right\rangle$ y dislikes x

(iii) someone x, everyone y (y dislikes x).

The contention is that "interpretively independent" quantifiers uniformly take widest branching, rather than first-order scope.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

However, this move cannot be generalized. One cannot account for the examples in (10) by treating 'any' as a wide-scope universal quantifier, whether or not one invokes branching structures:[4]

(10a) Not many soldiers have tortured any prisoners.
(b) Few soldiers have tortured any prisoners.
(c) I didn't say that Sam killed anyone.
(d) It is unlikely that any ticket will win.

Another distinguishing feature of interpretively independent elements is supposed to be their ability to coindex pronouns outside their own clauses and across sentences in discourse. But what does this amount to? If it is just the ability to serve as antecedents of pronouns in such environments, then type 2 quantifiers also have it:

(11a) Mary likes *a dog*, and *it* likes her.
(b) Mary found *a dog*. She told her mother that *it* was lost.

[4] (10a,b) state that the number of soldiers who have tortured even one prisoner is small. This is not the same as (i) [which is equivalent to (ii)], or (iii) [which is equivalent to (iv)].

(i) not $\begin{matrix}\text{many } x\text{: soldier } x \\ \text{every } y\text{: prisoner } y\end{matrix}$ x tortured y

(ii) $\sim$ (many x: soldier x) (every y: prisoner y) (x tortured y)

(iii) $\begin{matrix}\text{many } x\text{: soldier } x \\ \text{every } y\text{: prisoner } y\end{matrix}$ $\sim$ (x tortured y)

(iv) (many x; soldier x) (every y: prisoner y) $\sim$ (x tortured y)

(iii–iv) leave open the possibility that, although many soldiers are not torturers, many are; (10a,b) do not.

The interpretation of these sentences is also not given by (v):

(v) $\begin{matrix}\text{Few } x\text{: soldiers } x \\ \text{every } y\text{: prisoner } y\end{matrix}$ x tortured y

Since 'few soldiers' is monotone decreasing, whereas 'every prisoner' is monotone increasing, (v) has no coherent interpretation. [See Jon Barwise, "On Branching Quantifiers in English," *Journal of Philosophical Logic*, VIII, 1 (February 1979): 47–80.] By contrast, 'any' in (10a,b) can be analyzed as having the force of an existential quantifier within the scope of negation:

(vi) $\sim$ (many x: soldier x) (some y: prisoner y) (x tortured y)

The same point is illustrated by (10c,d), which have the interpretations (vii, viii) rather than (ix, x):

(vii) $\sim$ I said that [someone x (Sam killed x)]
(viii) Unlikely that [(some x: ticket x) (win x)]
(ix) everyone x $\sim$ I said that [Sam killed x]
(x) (every x: ticket x) unlikely that [win x]

In most lotteries each ticket is such that it is unlikely that it will win; however, it is seldom unlikely that there will be a winning ticket. Similarly, for each person, I may not have said that he or she was killed by Sam—even if I have said that Sam killed someone or other. (Contrast with analogous examples containing 'a certain'.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

One might argue that 'it' in (11) does not function as a bound variable, but rather is what Gareth Evans calls an "E-type pronoun."[5] But then one would have to show that pronouns with type 1 antecedents really do function as bound variables. For this, one would need a semantics, not provided by Hornstein, that distinguished variable binding from other types of pronoun/antecedent relations.[6]

The difficulty is especially pressing when it comes to definite descriptions. In chapter 3, Hornstein relies heavily on the fact that definite descriptions can serve as antecedents of pronouns in other clauses and in other sentences. Although some of the examples are candidates for a non-variable-binding analysis, he concludes without argument that they require definite descriptions to be of type 1.

This conclusion leads to the incorrect result that the descriptions in (12) unambiguously take widest interpretive scope:

(12a) If France is a monarchy and *the king of France* controls the army, then *he* is probably worried about Libya.
(b) Mitterand is not the king of France.
(c) Necessarily, the number of planets is greater than 7.
(d) John said that the largest prime is odd.
(e) John believed that the murderer of Jones was insane.
(f) The White House has always been occupied by the President of the United States.

Hornstein does not deny that sentences of this sort may be ambiguous. Rather, he maintains that these ambiguities do not involve

[5] "Pronouns, Quantifiers, and Relative Clauses (I)," in Mark Platts, ed., *Reference, Truth, and Reality* (Boston: Routledge & Kegan Paul, 1980). An *E-type pronoun* is a closed singular term whose reference is fixed by a description recoverable from the clause containing its antecedent. If 'it' in (11a) is such a pronoun, then the sentence expresses the conjunction of two independent propositions, rather than the existential proposition to the effect that the (conjunctive) property of being liked by Mary and liking her is possessed by at least one dog.

[6] Hornstein intends his theses about quantifier-pronoun coindexing to cover only non-E-type pronouns. However, he provides no semantics of cross-sentential indexing at all—let alone one demonstrating that pronouns with type 1 antecedents really function as bound variables, whereas those with type 2 antecedents do not. Prospects of doing this are not helped by the fact that some of Hornstein's examples of cross-sentential indexing of pronouns by type 1 quantifiers apparently cannot be genuine instances of variable binding. A case in point is example (6c) on page 44: "*Any citizen* can own a gun. In some countries *he* must." The truth of this discourse does not require "*x* can own a gun" and "In some countries *x* must own a gun" to be jointly true for each assignment of a citizen (of some contextually salient country) to '*x*'. Rather, what seems to be asserted is that every citizen (of a contextually salient country) can own a gun, and in some countries every citizen (of those countries) must own a gun. But, then, the pronoun in the example does not function as a variable bound by its antecedent. As a result, the alleged distinction between instances of cross-sentential indexing of pronouns by type 1 quantifiers and similar instances involving type 2 quantifiers remains mysterious.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

scope, do not result from "the language faculty," and should not be treated by a theory of logical form.

His argument rests on two claims. First, he maintains that familiar Russellian ambiguities "turn on" whether or not particular definite descriptions successfully refer—and whether or not an expression refers cannot be determined on the basis of linguistic competence alone (52). Second, he identifies large-scope Russellian readings with referential understandings of descriptions and small-scope readings with attributive understandings. Referential interpretations are said to be highly context-dependent and, hence, are claimed to go beyond the boundaries of semantics and "the language faculty" (53/4).

The upshot, for Hornstein, is that ambiguous Russellian sentences are assigned single "logical forms." The ambiguities, if they are to be represented at all, must be marked at some further level of mental representation. Here, Hornstein follows Chomsky in speculating that semantic interpretation may involve one set of rules, SI-1, internal to "the language faculty," and another entirely unspecified set of rules, SI-2, "due to the operation of other cognitive faculties that might perhaps be governed by relationships of the cognitive apparatus with facts about the world" (52). According to Hornstein, "aspects of interpretation that turn on whether or not a singular term refers to an actual object or to one that is believed to exist would be considered part of SI-2" (53). Nothing more is said about such rules.

There are several problems with this position. First, whether or not sentences containing descriptions are ambiguous is independent of whether or not the descriptions refer. These ambiguities can be detected using one's linguistic competence every bit as well as the ambiguities and other interpretive properties that Hornstein relies upon to support his analysis.[7] Second, Russellian ambiguities be-

[7] Although Hornstein's position encompasses all the sentences in (12), his argument focuses on (i):

(i) John didn't see the present King of France.

Here, Hornstein may have confused the question of whether (i) is ambiguous with the question of whether it has interpretations that differ in truth value. (The latter, but not the former, depends on the description's failing to refer.) Consideration of the range of examples in (12) should help in avoiding this confusion.

Hornstein extends the confusion to examples like (12e) and (13) below. About the ambiguity of the latter he says, "once again it concerns whether an expression ['a handsome Swede'] denotes actual existing objects. Interpretations that depend on such ontological concerns fall outside the theory of linguistic competence properly speaking" (55/6). In fact, the ambiguity does not depend on actually existing Swedes, and the interpretations can differ in truth value whether or not there are handsome Swedes. Of course, the truth of the wide-scope interpretation requires

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tween large- and small-scope interpretations of descriptions are not ambiguities between referential and attributive understandings.[8] Third, even if an interpretation is sensitive to context, that does not mean that it is nonsemantic—as the proliferation of context-sensitive semantics demonstrates. Fourth, if unspecified nonlinguistic rules of interpretation can save the theory from making false predictions about descriptions, then they can save it from making any predictions at all.

Hornstein uses this prediction-destroying device to deal with other problems, for example, the familiar scope ambiguity in (13):

(13) John believes that a handsome Swede loves Ingrid.

If his theory predicts anything, it predicts that (13) has only a small-scope interpretation. In reality, it also has a large-scope interpretation. Thus, Hornstein claims that the ambiguity is not properly linguistic, but rather is due to unspecified extralinguistic factors (SI-2 rules). Note the contrast. Here the "language faculty" is supposed to provide small-scope interpretations, and the extralinguistic wild card is invoked for large-scope interpretations; in the case of definite descriptions it is the other way around: the "language faculty" provides the allegedly nonsemantic, context-dependent large-scope interpretations, and the extralinguistic wild card is used for the small-scope interpretations. With such versatility the capacity to block predictions may be unlimited.

This presupposes, however, that predictions about interpretations would otherwise be forthcoming; whereas, strictly speaking, they would not. In motivating his account, Hornstein repeatedly appeals to data in which certain quantifiers are said to be unambiguously interpreted as having largest possible scope, while others are capable of taking different, clause-bounded, scopes. Indeed, a central aim of the theory is to predict the interpretive scopes of different quantifiers from their positions in psycho-logical syntax. But predictions

---

that at least one Swede be handsome; but so does the truth of (13′), whose linguistic credentials are not questioned:

(13′) John believes that a certain handsome Swede loves Ingrid.

Moreover, in chapter 4 Hornstein accepts an analysis according to which (13″) is linguistically ambiguous—despite the fact that its wide-scope interpretation is just the sort that is stigmatized as not properly linguistic in the case of (13).

(13″) John expects a handsome Swede to marry Ingrid.

[8] See Saul Kripke, "Speaker's Reference and Semantic Reference," in P. A. French, T. E. Uehling, Jr., and H. K. Wettstein, eds., *Contemporary Perspectives in the Philosophy of Language* (Minneapolis: Minnesota UP, 1979).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

about interpretive content require a theory of interpretive content, which Hornstein not only fails to provide but rejects in principle.[9]

In some cases this interpretive gap creates the illusion that a coherent analysis has been given, when none has. An example of this involves so-called 'donkey'-sentences, like those in (14):

(14a) Every farmer who owns a donkey beats it.
(b) Some farmer who owned a donkey beat it.

[9] The notion of interpretive scope, crucial to Hornstein's project, is essentially semantic. For example, to say that the italicized expressions in (i) have large scope is, in effect, to say that the rules for these expressions are applied first in determining the truth conditions of these sentences.

(i)(a) *S* (*eve~y x:Fx*) [true iff ⌜*Sy*⌝ is true of each *F*]
(b) ⌜~*S*⌝ (*the x:Fx*) [true iff there is a unique *F*-er and ⌜~*Sy*⌝ is true of it]
(c) *Not S* (*the x:Fx*) [true iff ⌜*S* (*the x:Fx*)⌝ is not true]

Hornstein, however, cannot give these characterizations, since he rejects the notion of truth they employ. Nor can he accept any alternative semantic characterization using the notions, *proposition, property, object, reference,* or *model,* since he also rejects these (in the sense in which they are standardly used in semantics). According to Hornstein, all these notions are dispensable in a proper theory of meaning; moreover, the interpretive properties of quantifiers, including their interpretive scopes, are said to be derivable from a purely syntactic theory of their structural characteristics in psycho-logical form.

There is a problem here, however, that Hornstein fails to acknowledge. A theory cannot make predictions about properties it does not countenance. What then is interpretive scope, if it is not the semantic phenomenon it has always been assumed to be? Although interpretive scope is supposed to be determined by syntactic scope (plus lexical properties) in psycho-logical form, the former is not to be identified with the latter, since various elements are said to have wide-interpretive scope even though they do not have wide-syntactic scope in psycho-logical form. Indeed, Hornstein thinks that interpretive scope is not properly scope at all—the latter being, on his view, a purely syntactic notion (26, 31, 50). Thus, he introduces a label, *interpretive independence,* to take the place of *wide-interpretive scope* (26, 31). But he never defines this term, and, in attempting to explain it, he appeals to Kripke's concept of rigid designation, despite the fact his theory makes no room for such a thoroughly semantic notion (31).

The situation, then, is this: Hornstein makes proposals that attempt to correlate alleged psycho-syntactic facts with well-known interpretive, semantic properties. As long as we retain our grasp on these properties we can evaluate Hornstein's proposals and test them for truth or falsity (where they are stated with enough precision to give clear results). When, however, semantic properties are rejected, with nothing put in their place, the notion of interpretive scope becomes empty, and the predictions Hornstein relies upon to motivate his theory evaporate.

Moreover, there is good reason to believe that any attempt to give content to interpretive properties like scope will reinstate the semantic notions Hornstein rejects. Roughly, for a sentence to have interpretive content is for it to represent the world as being a certain way; representing the world as being a certain way imposes conditions that must be satisfied if the world is to conform to the way it is represented to be. But these are just the truth conditions that Hornstein refuses to accept.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Problems posed by these sentences prompt Hornstein to claim that in (14a) 'a donkey' is a "quasi-generic" quantifier which is both of type 1 and of type 2.[10] This leads him to assign to (14a) a "logical form" consisting of a pair of representations: one in which 'a donkey' is a type 1 quantifier, and one in which it is a type 2 quantifier. By calling it type 1, Hornstein thinks he can explain how it can "bind" the pronoun outside the relative clause. By calling it type 2, he thinks he can explain why its scope is restricted to that clause (83/4). The two structures are supposed to "blend" in some unspecified way into a single interpretation, requiring every owner-donkey pair to be a beater-donkey pair (87/8). No importance is attached to the fact that type 1 and type 2 quantifiers have previously been associated with the incompatible properties of having, respectively, unrestricted and clause-bounded logical scope.

Since no explicit principles of interpretation are ever formulated, no explicit contradiction is forthcoming. By the same token, however, no explicit predictions about interpretations are made. At best, one can get a rough idea of the semantic facts Hornstein is trying to capture—keeping in mind that when things go wrong the rug may be pulled out, leaving it unclear how proposals about psycho-logical syntax are related to facts about interpretations.

In my opinion, these difficulties prevent Hornstein from doing justice to the semantic and psychological questions he raises. The class of data he surveys is rich, and the issues he discusses are worth investigating. He may even be right in thinking that such investigation will require theories of psycho-logical syntax. However, it will certainly require more.[11]

SCOTT SOAMES

Princeton University

[10] Hornstein provides no way of dealing with (14b).

[11] Space limitations have prevented me from commenting on Hornstein's criticism of representational, or truth-conditional, theories in semantics. The main force of the criticism rests, in my opinion, on an unwarranted demand that semantic theories play direct and central roles in explaining how semantic competence is acquired. For an extended discussion of the relationship between the two, see my "Semantics and Semantic Competence," in Stephen Schiffer and Susan Steele, eds., *The Second Arizona Colloquia on Cognitive Science* (Tucson: Arizona UP, forthcoming).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NEW BOOKS

HARDING, SANDRA: *The Science Question in Feminism.* Ithaca, N.Y.: Cornell, 1986. 271 p. Cloth $35, paper $9.95.

HARRELL, JEAN GABBART: *Soundtracks: A Study of Auditory Perception, Memory, and Valuation.* Buffalo, N.Y.: Prometheus Books, 1986. 117 p. $16.95.

HARTLE, ANN: *Death and the Disinterested Spectator: An Inquiry into the Nature of Philosophy.* Albany: SUNY Press, 1986. 263 p. Cloth $42.50, paper $14.95.

HARVEY, IRENE E.: *Derrida and the Economy of Différance.* Bloomington: Indiana UP, 1986. xv, 286 p. $24.95.

HAWORTH, LAWRENCE: *Autonomy: An Essay in Philosophical Psychology and Ethics.* New Haven, Conn.: Yale, 1986. vii, 248 p. $18.95.

HENDLEY, BRIAN: *Dewey, Russell, Whitehead: Philosophers as Educators.* Carbondale: Southern Illinois UP, 1986. xxi, 177 p. Cloth 19.95, paper $9.95.

HINSCH, WILFRIED: *Erfahrung und Selbstbewusstein.* Hamburg: Felix Meiner, 1986. x, 114 p. 36 DM.

HINTIKKA, MERRILL B., and JAAKKO HINTIKKA: *Investigating Wittgenstein.* New York: Basil Blackwell, 1986. xx, 326 p. $39.95.

HISKES, ANNE L., and RICHARD P. HISKES: *Science, Technology, and Policy Decisions.* Boulder, Colo.: Westview Press, 1986. ix, 198 p. Cloth $35, paper $15.95.

HOLLAND, JOHN H., KEITH HOLYOAK, RICHARD E. NISBETT, and PAUL R. THAGARD: *Induction: Processes of Inference, Learning, and Discovery.* Cambridge, Mass.: MIT Press, 1986. xvi, 385 p. $24.95.

HOLSCHER, LUDGER: *The Reality of the Mind.* New York: Methuen, 1987. viii, 341 p. $51.

HOLTON, GERALD: *The Advancement of Science, and Its Burdens.* New York: Cambridge, 1986. xi, 351 p. Cloth $39.50, paper $11.95.

HOURANI, GEORGE F.: *Reason and Tradition in Islamic Ethics.* New York: Cambridge, 1985. xv, 282 p. $39.50.

HUDSON, STEPHEN D.: *Human Character and Morality.* Boston, Mass.: Routledge & Kegan Paul, 1986. xvii, 164 p. $19.95.

HUTCHINSON, D. S.: *The Virtues of Aristotle.* New York: Methuen, 1987. ix, 139 p. $33.

IBBERSON, JOHN: *The Language of Decision.* Atlantic Highlands, N.J.: Humanities, 1986. vii, 156 p. $29.95.

IHDE, DON: *Consequences of Phenomenology.* Albany: SUNY Press, 1986. xi, 210 p. Cloth $34.50, paper $10.95.

INGARDEN, ROMAN: *The Work of Music and the Problem of Its Identity.* Berkeley: California UP, 1986. xvi, 181 p. $27.50.

JOHNSTON, DAVID: *The Rhetoric of Leviathan.* Princeton, N.J.: University Press, 1986. xx, 234 p. $25.

KAMMLER, CLEMENS: *Michel Foucault: Eine kritische Analyse seines Werks.* Bonn: Bouvier, 1986. 273 p. 68 DM.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

KATZ, JERROLD J.: *Cogitations.* New York: Oxford, 1986. 206 p. $24.95.

KATZ, MICHAEL J: *Templets and the Explanation of Complex Patterns.* New York: Cambridge, 1986. x, 128 p. $24.95.

KAVKA, GREGORY S.: *Hobbesian Moral and Political Theory.* Princeton, N.J.: University Press, 1986. xviii, 460 p. Cloth $45, paper $12.95.

KAYE, HOWARD L.: *The Social Meaning of Modern Biology.* New Haven, Conn.: Yale, 1986. ix, 184 p. $20.

KELLEY, DAVID: *The Evidence of the Senses.* Baton Rouge: Louisiana State UP, 1986. viii, 262 p. $27.50.

KIMBALL, BRUCE A.: *Orators and Philosopers: A History of the Idea of Liberal Education.* New York: Teachers College Press, 1986. xix, 293 p. $19.95.

KIRK, ROBERT: *Translation Determined.* New York: Oxford, 1986. xxiii, 276 p. $42.

KITCHENER, RICHARD F.: *Piaget's Theory of Knowledge: Genetic Epistemology and Scientific Reason.* New Haven, Conn.: Yale, 1986. ix, 230 p. $25.

KLEINIG, JOHN: *Ethical Issues in Psychosurgery.* Boston: Allen & Unwin, 1985. xv, 152 p. Cloth $19.95, paper $7.95.

KLOSKO, GEORGE: *The Development of Plato's Political Theory.* New York: Methuen, 1986. xviii, 263 p. Cloth $31.00, paper $11.50.

KOMERSAROFF, PAUL A.: *Objectivity, Science and Society: Interpreting Nature and Society in the Age of the Crisis of Science.* London: Routledge & Kegan Paul, 1986. ix, 454 p. $55.

KUNTZ, PAUL GRINDLEY: *Bertrand Russell.* Boston: Twayne, 1986. 186 p. $14.95.

KURTZ, PAUL: *The Transcendental Temptation: A Critique of Religion and the Paranormal.* Buffalo, N.Y.: Prometheus, 1986. xiv, 500 p. $19.95.

KVANVIG, JONATHAN L.: *The Possibility of an All-knowing God.* New York: St. Martin's, 1986. xv, 181 p. $27.50.

KVART, IGAL: *A Theory of Counterfactuals.* Indianapolis: Hackett, 1986. xxvii, 283 p. Cloth $39.50, paper $19.50.

LAND, STEPHEN K.: *The Philosophy of Language in Britain.* New York: AMS Press, 1986. 255 p. $37.50.

LANDMAN, FRED: *Towards a Theory of Information.* Dordrecht: Foris, 1986. xiv, 228 p. Cloth $34, paper $24.

LANGFORD, PETER: *Modern Philosophies of Human Nature.* Boston: Nijhoff, 1986. 265 p. $67.76.

LECLERC, IVOR: *The Philosophy of Nature.* Washington, D.C.: Catholic University Press of America, 1986. x, 220 p. $31.95.

LEE, KEEKOK: *A New Basis for Moral Philosophy.* London: Routledge & Kegan Paul, 1985. xv, 252 p. $36.95.

LEITES, EDMUND: *The Puritan Conscience and Modern Sexuality.* New Haven, Conn.: Yale, 1986. xi, 196 p. $17.50.

LEMOS, RAMON M.: *Rights, Goods, and Democracy.* Cranbury, N.J.: Delaware UP, 1986. 209 p. $28.50.

LEWIS, DAVID: *On the Plurality of Worlds.* New York: Basil Blackwell, 1986. ix, 276 p. Cloth $39.95, paper $12.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

LINGIS, ALPHONSO: *Libido: The French Existential Theories*. Bloomington: Indiana UP, 1986. xii, 128 p. $22.50.

LIPNER, JULIUS: *The Face of Truth*. Albany: SUNY Press, 1986. xii, 183 p. Cloth $44.50, paper $18.95.

LITTLE, DANIEL: *The Scientific Marx*. Minneapolis: Minnesota UP, 1986. xii, 244 p. Cloth $35.00, paper $14.95.

LIVINGSTON, ERIC: *The Ethnomethodological Foundations of Mathematics*. London: Routledge & Kegan Paul, 1986. xiii, 241 p. $49.95.

LOIZOU, ANDROS: *The Reality of Time*. Brookfield, Vt.: Gower, 1986. x, 206 p. $42.

LOMBARD, LAWRENCE BRIAN: *Events: A Metaphysical Study*. London: Routledge & Kegan Paul, 1986. xi, 272 p. $35.

LOWENTHAL, LEO: *Literature and the Image of Man: Communication in Society*, vol. 2. New Brunswick, N.J.: Transaction Books, 1986. vii, 344 p. $24.95.

LYONS, WILLIAM: *The Disappearance of Introspection*. Cambridge, Mass.: MIT Press, 1986. xiii, 193 p. $22.50.

MAC CORMAC, EARL R.: *A Cognitive Theory of Metaphor*. Cambridge, Mass.: MIT Press, 1986. x, 254 p. $22.50.

MACKEY, LOUIS: *Points of View: Readings of Kierkegaard*. Gainesville: University Presses of Florida, 1986. xxv, 205 p. $19.95.

MACNAMARA, JOHN: *A Border Dispute: The Place of Logic in Psychology*. Cambridge, Mass.: MIT Press, 1986, xv, 212 p. $22.50.

MALCOLM, NORMAN: *Nothing Is Hidden: Wittgenstein's Criticism of His Early Thought*. New York: Basil Blackwell, 1986. xii, 252 p. $34.95.

MARÍAS JULIÍAN: *The Structure of Society*. Tuscaloosa: Alabama UP, 1987. xvii, 224 p. $26.95.

MARKIE, PETER J.: *Descartes's Gambit*. Ithaca, N.Y.: Cornell, 1986. 278 p. $27.50.

MARTIN, MIKE W.: *Self-deception and Morality*. Lawrence: Kansas UP, 1986. x, 177 p. $19.95.

MATES, BENSON: *The Philosophy of Leibniz: Metaphysics and Language*. New York: Oxford, 1986. ix, 271 p. $29.95.

MATILAL, BIMAL KRISHNA: *Logic, Language and Reality: An Introduction to Indian Philosophical Studies*. Delhi: Motilal Banarsidass, 1985. xii, 447 p. 160 Rs.

MATILAL, BIMAL KRISHNA: *Perception: An Essay on Classical Indian Theories of Knowledge*. New York: Oxford, 1986. xv, 438 p. $65.

MATTICK, PAUL: *Social Knowledge*. Armonk, N.Y.: M. E. Sharpe, 1986. x, 137 p. $27.50.

MACCORMICK, NEIL, and OTA WEINBERGER: *An Institutional Theory of Law: New Approaches to Legal Positivism*. Boston: D. Reidel, 1986. xiv, 229 p. $43.

MAYO, BERNARD: *The Philosophy of Right and Wrong*. New York: Methuen, 1986. 176 p. Cloth $22.95, paper $12.95.

MCCARTHY, VINCENT A.: *Quest for a Philosophical Jesus*. Macon, Ga.: Mercer UP, 1986. xv, 240 p. $28.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

MCKUSICK, JAMES C.: *Coleridge's Philosophy of Language*. New Haven, Conn.: Yale, 1986. xiii, 175 p. $15.95.

MCDONALD, HENRY: *The Normative Basis of Culture*. Baton Rouge: Louisiana State UP, 1986. xii, 241 p. $25.

MCINERNY, RALPH: *Being and Predication: Thomistic Interpretations*. Washington, D.C.: Catholic University of America Press, 1986. xii, 319 p. $36.95.

MELVILLE, STEPHEN W.: *Philosophy beside Itself: On Deconstruction and Modernism*. Minneapolis: Minnesota UP, 1986. xxix, 188 p. Cloth $25, paper $10.95.

MEYER, MICHEL: *De la Problématologie: Philosophie, science et langage*. Brussels: Pierre Mardaga, 1986. 308 p. 229 FF.

MEYNELL, HUGO A.: *The Nature of Aesthetic Value*. Albany: SUNY Press, 1986. v, 158 p. $24.50.

MIDGLEY, MARY: *Evolution as a Religion*. New York: Methuen, 1986. 180 p. Cloth $33, paper $11.95.

MILLER, MITCHELL H., JR.: *Plato's Parmenides: The Conversion of the Soul*. Princeton, N.J.: University Press, 1986. xi, 299 p. $30.

MILNE, A. J. M.: *Human Rights and Human Diversity*. Albany: SUNY Press, 1986. viii, 186 p. Cloth $39.50, paper $12.95.

MINSON, JEFFREY: *Genealogies of Morals: Nietzsche, Foucault, Donzelot and the Eccentricity of Ethics*. New York: St. Martin's, 1985. x, 246 p. $25.

MORRIS, THOMAS V.: *The Logic of God Incarnate*. Ithaca, N.Y.: Cornell, 1986. 220 p. $19.95.

MUCKENHAUSEN, GABRIELE: *Wissenschaftstheorie und Kulturpessimismus*. Bonn: Bouvier, 1986. xi, 131 p. 48 DM.

MUNZ, PETER: *Our Knowledge of the Growth of Knowledge: Popper or Wittgenstein?* London: Routledge & Kegan Paul, 1986. xii, 353 p. $34.95.

MYERS, GERALD E.: *William James: His Life and Thought*. New Haven, Conn.: Yale, 1986. xxi, 628 p. $35.

NAGEL, THOMAS: *The View from Nowhere*. New York: Oxford, 1986. 272 p. $18.95.

NAKAMURA, HAJIME: *A Comparative History of Ideas*. London: Routledge & Kegan Paul, 1986. xx, 572 p. $58.95.

NATANSON, MAURICE: *Anonymity: A Study In the Philosophy of Alfred Schutz*. Bloomington: Indiana UP, 1986. xiv, 173 p. $25.

NELSON, R. J.: *The Logic of Mind*. Boston: D. Reidel, 1982. xvi, 392 p. $59.50.

NEWELL, R. W.: *Objectivity, Empiricism and Truth*. New York: Methuen, 1987. x, 124 p. $24.95.

NICHOLSON, LINDA J.: *Gender and History: The Limits of Social Theory in the Age of the Family*. New York: Columbia, 1986. x, 238 p. $27.50.

NOBO, JORGE LUIS: *Whitehead's Metaphysics of Extension and Solidarity*. Albany: SUNY Press, 1986. xv, 439 p. Cloth $49.50, paper $24.50.

NORTHROP, F. S. C.: *The Prolegomena to a 1985 Philosophiae Naturalis Principia Mathematica*. Woodbridge, Conn.: Ox Box Press, 1986. xvi, 73 p. $30.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NUSSBAUM, MARTHA C.: *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*. New York: Cambridge, 1986. xvi, 544 p. Cloth $59.50, paper $19.95.

O'HEAR, ANTHONY: *What Philosophy Is*. Atlantic Highlands, N.J.: Humanities, 1985. 316 p. Cloth $32, paper $9.95.

OESTERREICHER, JOHN M.: *The Unfinished Dialogue: Martin Buber and the Christian Way*. New York: Philosophical Library, 1986. 133 p. $14.95.

OLDENQUIST, ANDREW: *The Non-suicidal Society*. Bloomington: Indiana UP, 1986. xiii, 263 p. $19.95.

PAVEL, THOMAS G.: *Fictional Worlds*. Cambridge, Mass: Harvard, 1986. viii, 178 p. $20.

PEACOCKE, CHRISTOPHER: *Thoughts*. New York: Basil Blackwell, 1986. x, 175 p. $24.95.

PEARS, DAVID: *Ludwig Wittgenstein*. Cambridge, Mass.: Harvard, 1986. xxiv, 208 p. Paper $7.95.

PENELHUM, TERENCE: *Butler*. London: Routledge & Kegan Paul, 1986. x, 221 p. $39.95.

PENZ, G. PETER: *Consumer Sovereignty and Human Interests*. New York: Cambridge, 1986. xiv, 242 p. $34.50.

PERRY, THOMAS D.: *Professional Philosophy: What It Is and Why It Matters*. Boston: D. Reidel, 1986. xvii, 236 p. $39.50.

PHILP, MARK: *Godwin's Political Justice*. Ithaca, N.Y.: Cornell, 1986. x, 278 p. $29.95.

PIKE, SHIRLEY R.: *Marxism and Phenomenology*. Totowa, N.J.: Barnes & Noble, 1986. 201 p. $28.50.

PINCOFFS, EDMUND L.: *Quandaries and Virtues: Against Reductivism in Ethics*. Lawrence: Kansas UP, 1986. x, 186 p. $19.95.

PIVČEVIĆ, EDO: *The Concept of Reality*. New York: St. Martin's, 1986. ix, 296 p. $30.

POJMAN, LOUIS P.: *Religious Belief and the Will*. London: Routledge & Kegan Paul, 1986. xiii, 258 p. $32.50.

POLKA, BRAYTON: *The Dialectic of Biblical Critique: Interpretation and Existence*. New York: St. Martin's, 1985. x, 191 p. $25.

POPE, WHITNEY: *Alexis de Tocqueville: His Social and Political Theory*. Beverly Hills, Calif.: Sage, 1986. 159 p. Cloth $16.95, paper $8.95.

PRADO, C. G.: *Rethinking How We Age: A New View of the Aging Mind*. Westport, Conn.: Greenwood Press, 1986. xi, 174 p. $29.95.

RACHELS, JAMES: *The End of Life*. New York: Oxford, 1986. vii, 204 p. $17.95.

RICHARDS, DAVID A. J.: *Toleration and the Constitution*. New York: Oxford, 1986. xvii, 348 p. $29.95.

RIFKIN, JEREMY: *Declaration of a Heretic*. Boston: Routledge & Kegan Paul, 1985. x, 140 p. Cloth $22, paper $7.95.

ROBERTS, ROBERT C.: *Faith, Reason, and History*. Macon, Ga.: Mercer UP, 1986. vi, 145 p. $22.50.

ROCKMORE, TOM: *Hegel's Circular Epistemology*. Bloomington: Indiana UP, 1986. xv, 202 p. $22.50.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

RODERICK, RICK: *Habermas and the Foundations of Critical Theory.* New York: St. Martin's, 1986. xi, 194 p. Cloth $37.50, paper $12.95.

RÖHR, FERDINAND: *Die pädagogische Theorie im Denken von Karl Jaspers.* Bonn: Bouvier, 1986. x, 310. 88 DM.

ROLSTON III, HOLMES: *Philosophy Gone Wild.* Buffalo, N.Y.: Prometheus, 1986. 269 p. $19.95.

ROSENBAUM, ALAN S.: *Coercion and Autonomy: Philosophical Foundations, Issues, and Practices.* Westport, Conn.: Greenwood Press, 1986. xii, 196 p. $32.95.

ROSENBERG, JAY F.: *The Thinking Self.* Philadelphia: Temple UP, 1986. ix, 267 p. $37.95.

ROSENTHAL, SANDRA B.: *Speculative Pragmatism.* Amherst: Massachusetts UP, 1986. 213 p. $25.

ROY, RAMASHRAY: *Self and Society: A Study in Gandhian Thought.* Beverly Hills: Sage, 1985. 205 p. $25.

RUSE, MICHAEL: *Taking Darwin Seriously.* New York: Basil Blackwell, 1986. xv, 303 p. $24.95.

SALLIS, JOHN: *Delimitations: Phenomenology and the End of Metaphysics.* Bloomington: Indiana UP, 1986. xi, 210 p. $22.50.

SCHEFFLER, ISRAEL: *Inquiries: Philosophical Studies of Language, Science, and Learning.* Indianapolis: Hackett, 1986. xv, 400 p. Cloth $30, paper $19.50.

SCHIER, FLINT: *Deeper into Pictures: An Essay on Pictorial Representation.* New York: Cambridge, 1986. ix, 226 p. $29.95.

SCHLAGEL, RICHARD H.: *Contextual Realism: A Meta-physical Framework for Modern Science.* Buffalo, N.Y.: Paragon House, 1986. 308 p. $21.95.

SCHRAG, CALVIN O.: *Communicative Praxis and the Space of Subjectivity.* Bloomington: Indiana UP, 1986. ix, 219 p. $27.50.

SCHÜLLER, BRUNO: *Wholly Human: Essays on the Theory and Language of Morality.* Washington, D.C.: Georgetown UP, 1986. 220 p. Cloth $17.95, paper $9.95.

SCHWARTZ, THOMAS: *The Logic of Collective Choice.* New York: Columbia, 1986. xiv, 315 p. $35.

SEESKIN, KENNETH: *Dialogue and Discovery: A Study in Socratic Method.* Albany: SUNY Press, 1987. viii, 179 p. Cloth $39.50, paper $12.95.

SEIDLER, VICTOR J.: *Kant, Respect and Injustice: The Limits of Liberal Moral Theory.* London: Routledge & Kegan Paul, 1986. xii, 244 p. $35.

SHAPIRO, IAN: *The Evolution of Rights in Liberal Theory.* New York: Cambridge, 1986. x, 326 p. Cloth $39.50, paper $11.95.

SHELP, EARL E.: *Born to Die? Deciding the Fate of Critically Ill Newborns.* New York: Free Press, 1986. xiii, 250 p. $19.95.

SHEROVER-MARCUSE, ERICA: *Emancipation and Consciousness.* New York: Basil Blackwell, 1986. viii, 211 p. $24.95.

SHOEMAKER, SYDNEY, and RICHARD SWINBURNE: *Personal Identity: Great Debates in Philosophy.* New York: Basil Blackwell, 1984. 158 p. $24.95.

SIEBERT, RUDOLF J.: *The Critical Theory of Religion: The Frankfurt School.* Berlin: Mouton, 1985. xvi, 722 p. $85.90.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SLEEPER, R. W.: *The Necessity of Pragmatism: John Dewey's Conception of Philosophy.* New Haven, Conn.: Yale, 1986. xii, 236 p. $21.50.

SMITH, LAURENCE D.: *Behaviorism and Logical Positivism: A Reassessment of the Alliance.* Stanford, Calif.: University Press, 1986. ix, 398 p. $42.50.

SMITH, PETER, and O. R. JONES: *The Philosophy of Mind.* New York: Cambridge, 1986. xviii, 282 p. Cloth $34.50, paper $8.95.

SMITH, QUENTIN: *The Felt Meanings of the World: A Metaphysics of Feeling.* West Lafeyette, Ind.: Purdue UP, 1986. xv, 352 p. $19.75.

SOBLE, ALAN: *Pornography: Marxism, Feminism, and the Future of Sexuality.* New Haven, Conn.: Yale, 1986. x, 202 p. $21.50.

SPERBER, DAN, and DEIRDRE WILSON: *Relevance: Communication and Cognition.* Cambridge, Mass.: Harvard, 1986. viii, 279 p. Cloth $25, paper $8.95

SPITZ, ELLEN HANDLER: *Art and Psyche: A Study in Psychoanalysis and Aesthetics.* New Haven, Conn.: Yale, 1985. 188 p. $18.95.

STAMBAUGH, JOAN: *The Real Is Not the Rational.* Albany: SUNY Press, 1986. xi, 130 p. Cloth $29.50, paper $9.95.

STOVE, D. C.: *The Rationality of Induction.* New York: Oxford, 1986. 231 p. $37.

SUCHTING, W. A.: *Marx and Philosophy.* New York: NYU Press, 1986. xiii, 133 p. $35.

SWARTZ, NORMAN: *The Concept of Physical Law.* New York: Cambridge, 1985. xi, 220 p. $29.95.

SWINBURNE, RICHARD: *The Evolution of the Soul.* New York: Oxford, 1986. vi, 323 p. $45.

TAMNY, MARTIN, and K. D. IRANI: *Rationality in Thought and Action.* Westport, Conn.: Greenwood Press, 1986. xx, 286 p. $37.50.

TAYLOR, PAUL W.: *Respect for Nature: A Theory of Environmental Ethics.* Princeton, N.J.: University Press, 1986. ix, 329 p. Cloth $37.50, paper $12.50.

THOMAS, WILLIAM: *Mill.* New York: Oxford, 1985. 134 p. $13.95.

THOMPSON, JUDITH JARVIS: *Rights, Restitution, and Risk: Essays in Moral Theory.* William Parent, ed. Cambridge, Mass.: Harvard, 1986. x, 269 p. Cloth $29.95, paper $9.95.

TUCKER, D. F. B.: *Law, Liberalism, and Free Speech.* Totowa, N.J.: Rowman & Littlefield, 1986. viii, 212 p. Cloth $32.50, paper $13.50.

TURNER, WILLIAM B.: *Nothing and Nonexistence: The Transcendence of Science.* New York: Philosophical Library, 1986. xxiii, 431 p. $27.50.

VANDEVEER, DONALD: *Paternalistic Intervention: The Moral Bounds of Benevolence.* Princeton, N.J.: University Press, 1986. xii, 452 p. $40.

VAN LEER, DAVID: *Emerson's Epistemology: The Argument of the Essays.* New York: Cambridge, 1986. xviii, 282 p. $29.95.

VAUX, KENNETH: *Powers that Make Us Human: The Foundations of Medical Ethics.* Champaign: Illinois UP, 1986. 145 p. $16.95.

VEATCH, HENRY B.: *Human Rights: Fact or Fancy?* Baton Rouge: Louisiana State UP, 1985. xi, 258 p. $30.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

VUILLEMIN, JULES: *What Are Philosophical Systems?* New York: Cambridge, 1986. ix, 163 p. $39.50.

WALTON, DOUGLAS N.: *Courage: A Philosophical Investigation.* Berkeley: California UP, 1986. xiv, 241 p. $25.

WANG, HAO: *Beyond Analytic Philosophy.* Cambridge, Mass.: MIT Press, 1986. xii, 273 p. $17.50.

WASHINGTON, JOHNNY: *Alain Locke and Philosophy: A Quest for Cultural Pluralism.* New York: Greenwood Press, 1986. xxxii, 246 p. $35.

WEISS, PAUL: *Toward a Perfected State.* Albany: SUNY Press, 1986. xii, 459 p. Cloth $44.50, paper $18.95.

WERNER, JÜRGEN: *Darstellung als Kritik: Hegels Frage nach dem Anfang der Wissenschaft.* Bonn: Bouvier, 1986. 215 p. 58 DM.

WHITE, DAVID A.: *Logic and Ontology in Heidegger.* Columbus: Ohio State UP, 1986. xiv, 245 p. $22.50.

WHITE, MICHAEL J.: *Agency and Integrality.* Boston: D. Reidel, 1985. xiii, 283 p. $44.

WHITE, MORTON: *Philosophy, The Federalist, and the Constitution.* New York: Oxford, 1987. xi, 273 p. $29.95.

WHITNEY, CHARLES: *Francis Bacon and Modernity.* New Haven, Conn.: Yale, 1986. x, 234 p. $18.50.

WILSON, FRED: *Laws and Other Worlds.* Boston: D. Reidel, 1986. xv, 328 p. $69.

WILSON, GEORGE M.: *Narration in Light: Studies in Cinematic Point of View.* Baltimore: Johns Hopkins UP, 1986. x, 223 p. $22.50.

WIND, EDGAR: *Hume and the Heroic Portrait.* New York: Oxford, 1986. xv, 140 p. $62.

WOODHOUSE, C. M.: *Gemistos Plethon: The Last of the Hellenes.* New York: Oxford, 1986. xxi, 391 p. $74.

YACK, BERNARD: *The Longing for Total Revolution: Philosophic Sources of Social Discontent from Rousseau to Marx and Nietzsche.* Princeton, N.J.: University Press, 1986. xvii, 385 p. $35.

YURKOVETS, IVAN: *The Philosophy of Dialectical Materialism.* L. LEZHNEVA and A. ZUR, trans. Moscow: Progress Publishers, 1984. 268 p. $7.95.

ZANDVOORT, HENK: *Models of Scientific Development and the Case of Nuclear Magnetic Resonance.* Boston: D. Reidel, 1986. xiii, 305 p. $49.50

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The Journal of Value Inquiry

Executive Editor
James Wilbur

The Journal of Value Inquiry is an international philosophical quarterly devoted to the stimulation and communication of current research in axiology. Papers may address themselves to questions concerning the nature, origin, experience and scope of value in general, as well as to more restricted studies concerning problems of values in ethics, social and legal theory and practice, science, aesthetics, and methodology. The journal is committed to no position or program and, in consequence, invites papers without restriction of method, scope, or orientation.

Subscription Information **ISSN** 0022–5363
1987, Volume 21 (4 issues)
Institutional rate: Dfl. 181.00/US$76.00 incl. postage/handling
Private rate: Dfl. 75.00/US$31.50 incl. postage/handling

Private subscriptions should be sent direct to the publishers

Back Volume(s) Available
Volumes 1–20 (1967–1986)

Price per Volume
excl. postage
Dfl. 145.00/US$50.00

Martinus Nijhoff Publishers

Kluwer academic publishers group

P.O. Box 989
3300 AZ Dordrecht
The Netherlands

101 Philip Drive
Norwell,
MA 02061, U.S.A.

Falcon House
Queen Square
Lancaster, LA1 1RN, U.K.

PRINCETON • UNIVERSITY • PRESS

Socrates and the State
Richard Kraut

Richard Kraut shows that Plato explains Socrates' refusal to escape from jail and his acceptance of the death penalty as arising not from a philosophy that requires blind obedience to every legal command but from a highly balanced compromise between the state and the citizen.

"Richard Kraut's important and impressive book is the best available discussion of the *Crito* and of Socrates' political views and, in several ways, the best available book on Socrates."
—*T. H. Irwin, Ethics*

Now available in paperback
Paper: $10.50. Cloth: $45.00

Plato's Late Ontology
A Riddle Resolved
Kenneth M. Sayre

Kenneth M. Sayre proposes new interpretations of the *Philebus* and the *Parmenides* that link the contents of these dialogues with other longstanding puzzles of Platonic scholarship.

"Sayre is to be commended for the fine historical and exegetical work he has done, particularly as it pertains to the second part of the *Parmenides*. His unique and illuminating discussion of this extremely difficult text is in itself well worth the price of the book."
—*Daryl Pullman, Eidos*

Now available in paperback
Paper: $16.50. Cloth: $39.50

At your bookstore or
41 William Street **Princeton University Press** Princeton, NJ 08540

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIV, NO. 9, SEPTEMBER 1987

FREEDOM AND VALUE

I will argue in this essay for an unorthodox necessary condition of free action. I believe that this condition will be an important component of any adequate theory of freedom.

I

In order to begin to construct a theory of free action we need, at the very least, an intuition and a question. Without some initial, relatively fixed beliefs about free action to start with, we would not know what we are theorizing about. And, apart from questions concerning the subject of our theory, we would not have any motivation to theorize; nor would we have much of a clue about how to begin. Part of the reason why free action has been the occasion for such controversy and confusion is that philosophers have not worked hard enough to display the intuitive origins and motivating questions for their favored theories of free action.

The main intuition that my discussion relies on is that, among all our actions, our free actions are fully, or most truly, our own. Our free actions do not merely belong to us as any of our actions must. They belong to us in a particularly intimate way. Obviously these expressions of the intuition can be taken in many senses. But, if there is a coherent concept of free action which typically governs our use of phrases like 'acted freely' in utterances such as "She should not be blamed unless she acted freely," then I claim that there is a sense in which we all would agree that our free actions are fully, or most truly, our own, and that, understood in whatever sense this is, this is one of our firmest and most widely shared convictions about free action.

Judging from the recent literature, this appears to be an intuition that guides most theorizing on the subject. To cite only a few prominent examples: in discussing the unfreedom of a person unwillingly addicted to a drug, Harry Frankfurt[1] writes,

[1] "Freedom of the Will and the Concept of a Person," this JOURNAL, LXVIII, 1 (Jan. 14, 1971): 5–20.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

0022-362X/87/8409/0465$02.20 © 1987 The Journal of Philosophy, Inc.

Digitized by Arya Samaj Foundation Chennai and eGangotri

> The unwilling addict identifies himself . . . through the formation of a second-order volition, with one rather than with the other of his conflicting first-order desires. He makes one of them *more truly his own* and, in so doing, he withdraws himself from the other. It is in virtue of this identification and withdrawal . . . that the unwilling addict may meaningfully make the analytically puzzling statements that the force moving him to take the drug is a force *other than his own,* and that it is not of his own free will but rather against his will that this force moves him to take it (13; my emphasis).

Wright Neely[2] says in a similar vein, "we must find a way of saying that some desires are more intimately related to the self—are *more its own*—than others without adopting the 'solution' . . . of simply putting another man inside the man" (43; my emphasis). In a critique of Frankfurt's and Neely's accounts, Michael Slote[3] has said that acting of one's own free will "involves something more than willingness taken together with our previous rationality conditions. And I believe that this further condition, in simplest terms, comes down to the requirement that our actions be fully *our own*" (150). Slote continues,

> the assumption that acting of one's own free will requires that one's actions be fully one's own also helps to explain what the words 'one's own' are doing in the phrase 'of one's own free will'. . . . In order to act of one's *own* free will, then, one's action must be fully one's own; one must in some sense act autonomously (*ibid.*).

Again, there is a legitimate fear that, by emphasizing such similarity in phraseology, especially when the phrase seems so open to divergent interpretation, broad acceptance of the initial intuition becomes trivial. I will try to dispel this fear below, indicating how the idea that free acts are fully our own may be used to illuminate other settled intuitions about freedom, such as the belief that we act freely only when we have control over whether or not we act as we do.

The question that motivates my theoretical proposals about freedom is this: What general abilities must a person have, beyond whatever abilities agency requires, in order to act freely? This question seems to have informed most theories, since nearly all have maintained—though nearly always without defense—that we act freely just when we have the ability to do otherwise than we do. Curiously, the question has so dominated the subject that, for many, it has ceased to be a live question at all. Its answer (they say) is uninterest-

[2] "Freedom and Desire," *Philosophical Review,* LXXXIII, 1 (January 1974): 32–54.

[3] "Understanding Free Will," this JOURNAL, LXXVII, 3 (March 1980): 136–151.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ingly obvious—the ability to do otherwise. The real question has been thought to be: What limitations, if any, does ability to do otherwise put on possible causal histories of our behavior, and do such restrictions ever actually obtain? The problem of freedom has been the problem of freedom *and* determinism.

One of the drawbacks of this orthodox approach is that it is too rash. Having the ability to do otherwise, in an apparently uncontroversial sense, does not suffice for free action. If I cannot anticipate an immediate consequence of one of my free acts, so that I bring it about unintentionally, then I do not freely bring about that consequence. I do not freely perform the action described in terms of that consequence (spilling the coffee, stepping on a thumbtack, hurting a friend's feelings). Yet I may well have the ability to do otherwise, since I may have the ability to do otherwise than perform the action that generated the unintended effect.

Having the ability to do otherwise in the relevant sense does not distinguish free agents, because it is a necessary characteristic of anyone who does anything intentionally, and intentionality is insufficient for freedom. Briefly, one has the ability to act when one has the power to translate one's will into successful action, other things being equal (i.e., where one has the opportunity to perform the action, one does not change one's mind, one makes a good effort if effort is needed, and so forth). Anyone who does something intentionally has the ability to act at that time, since the action's intentionality means that the behavior resulted from one's will in the appropriate way. The pertinent ability to translate one's will into successful action is nothing other than the ability to act intentionally.[4]

Since acting intentionally suffices for having the ability to act as one did, it also suffices for having the ability to refrain from what one did. For, if one's actual acting depended in the right way on one's will, then one's will to *refrain* from that action would have been enough to bring about intentional refraining (under similar *ceteris paribus* conditions). If my acting owes to my willing, then nothing could stand in the way of my refraining from that act should I have willed to refrain from it.[5] This claim may seem utterly implausible for the garden variety of psychologically compelled action, like acting

[4] I call this sense of ability "ordinary ability." I discuss this and other senses of ability more carefully in "Ordinary Ability and Free Action," *Canadian Journal of Philosophy*, XVII, 2 (June 1987). My present claims about intentional action and ability to do otherwise are also defended more thoroughly in that article.

[5] I believe that this holds true even where the actual performance is overdetermined. Further discussion is not possible here.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

out of drug addiction, irresistible obsession, or hypnotic inducement. But this worry stems from a confusion. Psychological compulsion typically undermines freedom while preserving intentionality. The addict's drug use is the intentional result of her irresistible will to take the drug. Since the addict uses the drug because she wants to (even though she may not ultimately value either the want or the behavior), her having the will not to use the drug instead of her present desire would yield the appropriate refraining. The addiction utilizes the addict's ability to take the drug willfully. That ability cannot be present without the complementary ability to act successfully on a resolute will to refrain.

The addict's problem is that, although her behavior originates with her intentions, she is not in *control* of the intentions that move her to take the drug and so is not in control over whether or not she uses the drug. The addict has the bare ability to do otherwise; or else her drug use would be involuntary, "mere behavior." What she lacks is control over the exercise of this ability. She acts on a will she cannot govern. Her drug use is hers because it is intentional; it is not fully or most truly hers, because she lacks control over whether or not she engages in it.[6]

It will likely be objected that this treatment of the ability to do otherwise cannot be an appropriate response to the orthodox tenet that free agency consists in having that ability. Certainly many have thought that this ability is no trivial thing, as freedom is no trivial thing. But, if persons as unfree as addicts may have this ability, then (the objection runs) the important notion of ability has been passed over and a largely insignificant concept of ability put in its place. If my contention is that ability to do otherwise does not reveal anything special about free action since it is a prerequisite of any intentional action, thus necessary for agency itself whether free or unfree, then the point may be merely verbal. Perhaps the traditional debate over the causal implications of avoidability is really just a debate over what control amounts to. After all, even if the addict has the minimal ability to do otherwise, does she not still lack the ability to do otherwise *in the face* of her desire for the drug?

There are three responses to this worry. First, if traditional claims about freedom and ability to do otherwise are really claims about what I term "control," then, contrary to the tradition, the primary question about the abilities distinctively required for free agency is not only a live question, but one which has not been answered satis-

[6] I explain below why lack of control would make the addict's action less than fully her own even if she relishes her addiction.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

factorily. Holding that control over whether or not we act as we do is what matters for freedom, far from answering our question, raises it all over again: What general abilities distinguish agents who have this control? The problem of freedom remains much more than the problem of freedom and determinism, for we still do not understand what abilities freedom involves, much less their implications for determinism.

Second, there is good reason to employ the 'ability to do otherwise' terminology to refer to that power which any intentional agent possesses, reserving 'control' for the more specialized power of free agents. When ability to do otherwise is used perspicuously, it suggests an instance of the ability to act intentionally. Once this connection is made, the ability to do otherwise is naturally seen to consist in some relationship between willing and acting which will be incapable of accounting for defects in agents' relations to their own wills (as in cases of psychological compulsion). Unless 'ability' is going to mean much more than we ordinarily use it to mean, the constitutive features of free agency should not be lumped together under the rubric "ability to do otherwise."[7]

The third reason why the issue I have brought before traditional theories of free action is not trivial is that, even once we adopt the terminology of control or set aside a special use of 'ability to do otherwise', it still is not the case that having control, or this specialized ability to do otherwise, exhausts the content of free agency. The central point of this essay is to argue that, however we understand control, there is a further, equally significant ability necessary for free action. This is an ability to criticize courses of action competently by relevant normative standards. This ability lends normative substance to the idea of free action, for it entails that full freedom of action is impossible without a certain appreciation of values.

The belief that free actions are those which are fully our own has not played any significant role in the argument thus far. I have affirmed that one central condition of freedom is control. Lacking control over our actions is one way in which they can be less than fully our own. As the remarks by Frankfurt and Neely cited above indicate, we can find it helpful to express in terms of their limited ownership of their actions the unfreedom of those who act out of

[7] There is admittedly some unclarity in the terminology of control as well. Being in control of one's conduct can mean that the behavior is habitual or well entrenched, the opposite of impulsive or unpredictable. This has nothing especially to do with freedom. I use 'control' to indicate our control over whether or not we perform a particular action, regardless of that action's being performed in a controlled way.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

compulsive motives. This is not yet an argument that control is necessary for freedom. The image of ownership must be clarified first. At best, it gives us a small clue about how to understand the place of control in freedom. For the moment, note that the ownership intuition suggests that free action may not be an all-or-nothing matter. Actions can be our own to a greater or lesser extent. If an action that is fully our own is free unqualifiedly, then perhaps an action that is less than fully ours can be free to some lesser degree. In envisaging an element of freedom beyond control, then, I am not claiming that persons who have control over their actions may be no more free than those who act out of compulsion. I shall argue only that their actions may be less than fully free.

II

The idea that free action has a normative component, and thus is of direct concern for value theory as well as for metaphysics, has been developed in different ways in two recent and significant essays. My own position will be more intelligible if we look briefly at some of the related claims these papers advance.

The first paper is Gary Watson's "Free Agency."[8] Watson argues that views, like Frankfurt's, which try to comprehend freedom solely in terms of structural features of the will cannot adequately explain the identification of persons with their free actions, until they attend to the place of values in the motivational structure of free action. Watson contends that free agents are set apart by having the ability to do what they most want, or value, all things considered, in the circumstances they face. Whereas intentional action depends on the ability to translate one's will into appropriate behavior, freedom (according to Watson) consists in the further ability to translate one's value judgments into effectual willing, without interference from motives that arise independently of one's values. Elsewhere Watson draws the broader consequence of this for theories of freedom:

> Apparently, the belief in free will presumes for human beings a special and puzzling status in the natural world, one that is central to our moral outlook. A convincing treatment of the problem of free will must give an account of the values in question, as well as of this presumed status. For this reason, free will is as much a problem in moral philosophy as in metaphysics. Our view of what free will is depends on our view of what matters (*Free Will,* Introduction, p. 2).

[8] This JOURNAL, LXXII, 8 (April 24, 1975): 205–220. Reprinted in Watson's anthology *Free Will* (New York: Oxford, 1982), pp. 96–110. Page references will be to the former.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Watson's view seems to do rather well as an account of the control that free agents must possess.[9] Watson notes that being able to do what one wants is not distinctive of freedom, because anyone who acts intentionally has this ability (205). In attempting to distinguish free action from merely intentional action on the grounds that a free agent's will is open to the governance of her values, Watson takes a position close to the view on psychological compulsion I sketched above. I argued that, because persons who act out of some psychological compulsion must have the abilities to make willing effective in action which any agent has (abilities which necessarily include ability to do otherwise), the special problem that compulsive motivation poses for freedom concerns the agent's inability to govern her will. Thus, Watson plausibly seeks to locate the difference between freedom and compulsion in the sensitivity of the will to some special source of motivation—in this case values—on which a free agent acts. Valuation seems an especially good candidate for this role, since most cases of unfreedom due to lack of control are cases in which the agent either is unable to do what she thinks best or would have been unable to act on her values had she judged it best to do otherwise. Where I do not know what I am doing or do not mean to do it, lack of intentionality blocks the potential efficacy of my values. Where I do what I am psychologically compelled to do, helplessness with respect to my will blocks the translation of values into action.

My strongest differences with Watson come on two related points. First, Watson seems to think that control is sufficient for freedom. Watson assumes that lack of intention and psychological compulsion ("obstruction by one's own will") are the only sorts of threat to free agency. This leaves no room for the kinds of cases I examine below, in which persons whose conduct is open to the rule of whatever values they might accept nonetheless act unfreely to the extent that they either are not aware of the values that actually apply to their actions or do not appreciate those values well enough. Watson neglects the diminished freedom of the young child or of the person whose moral education was severely deficient.

As a result of Watson's failure to see that freedom requires more than control, he also does not grasp the extent to which a conception of value needs to be drawn into an account of free action. A person

[9] Serious arguments to the contrary have been offered by Christopher S. Hill in "Watsonian Freedom and Freedom of the Will," *Australasian Journal of Philosophy*, LXII, 3 (September 1984): 294–298. Since my present aim is not to analyze control, I will not deal with Hill's objections here.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

who suffers from a severely deprived upbringing may be able to form evaluative judgments and act on them in the way Watson's theory specifies. But this person's actions will be less than fully her own if, for instance, she forms these judgments in deviant ways, being understandably ignorant or unappreciative of the standards of conduct others endorse. Contrary to Watson, I want to argue that freedom is a normative notion to the extent that attributions of free action in any particular context depend for their correctness partly on the *content* of the agent's normative understanding, not just on the agent's having some valuational point of view or other.

Watson's omission of the evaluative substance of freedom is substantially remedied in Susan Wolf's intriguing paper "Asymmetrical Freedom."[10] Wolf also aims to draw the problem of free action further into the arena of moral philosophy. She believes that its treatment as a purely metaphysical issue may be responsible for its apparent intractability (160/1). Unlike Watson, though, Wolf thinks that being able to guide one's conduct by means of the values one accepts is not a strong enough condition of free agency. Freedom, on her account, demands that one hold the values one ought to hold. Quite in line with my differences with Watson, Wolf believes that obstacles to competent appreciation of the norms that apply to our actions are as much impediments to full freedom as are certain obstacles to the expression of our evaluative judgments in our will or certain obstacles to the realization of our will in our conduct.

In defending this point, Wolf takes up the example of someone who deliberately performs criminal acts but who cannot be expected to appreciate their wrongness because his upbringing was seriously deprived. Wolf argues that a deprived childhood diminishes this person's freedom, despite the fact that the internal connections between his judgments, decisions, and behavior are free of abnormal interference. She writes,

> . . . he, unlike us, could not have had reasons even though there were reasons around. The problem is not that his reason was functioning improperly, but that his data were unfortuitously selective. Since the world for him was not suitably cooperative, his reason cannot attain its appropriate goal (160).

Wolf comes closer than any other recent writer to illuminating the normative dimension of free action. But her outline of an account of freedom labors under some worrisome difficulties. First, she be-

[10] This JOURNAL, LXXVII, 3 (March 1980): 151–166.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

lieves, wrongly I think, that revealing the evaluative character of freedom also helps resolve the question of freedom and determinism. Wolf contends that, if freedom is the freedom to do what there are good and sufficient reasons to do, then an agent who does what is best may (though need not) be psychologically determined to have done so—determined by the action's being best, that is—without being at all unfree. The reverse is true for someone who does what is wrong. The determination of wrong action must upset freedom, Wolf contends, because it prevents an agent from being able to do what is really best. Wolf concludes that both the compatibility of freedom and determinism and the necessity for free agency of the ability to do otherwise are contingent on the moral quality of the agent's conduct.

This argument fails at a number of places. Wolf offers no defense whatsoever for the contention that many cases of right action are instances of determined action. I grant that we can speak of a good person's actions as being in some sense determined by their moral desirability. This shows nothing, however, about the compatibility of freedom and determinism, for what is in dispute in the traditional problem of free will is *causal* determination. As far as I can tell, no one finds a person's freedom to be threatened simply by the sensitivity of the person's behavior to the values that apply to it. The traditional philosophical puzzle concerns *how* values guide the person's life. If the life is the causal product of circumstance, upbringing, and heredity, and these are not within the person's control, then freedom seems absent (so long as one has incompatibilist sentiments or can somehow conjure them up). Even if, as Wolf says (152/3), acting contrary to our interests is normally a sign of insanity and being moved to act without regard to the world outside us would make our lives capricious in a dangerous way, this does nothing to establish that the normal dependence of action on interests and of interests on environment, which Wolf labels "psychological determinism" (152), is causal dependence. As it stands, Wolf's account is irrelevant to the debate between compatibilists and incompatibilists.

Wolf's confusion about determinism becomes evident when, at the end of the essay, she argues that an undetermined act need not be a mere product of chance, since undetermined decisions normally are not random (165/6). But, if we accepted her earlier point that indeterminism holds out only the possibilities of insanity or whim, then we should not buy this argument. And, if we accept this argument, we should not be ready to admit that good persons are normally determined to act as they do. Why can they not just (indeterministically) decide, on the best of grounds, to do what is best?

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Wolf's claim that ability to do otherwise is necessary for freedom only where there is good and sufficient reason to do otherwise is similarly weak. She relies heavily on familiar remarks like "He couldn't hurt a fly" for evidence that admirable conduct may remain praiseworthy, hence free, even in the absence of ability to do otherwise (156). Such remarks, taken literally, do not withstand close scrutiny. Wolf must grant that the best sort of person would hurt someone if there were sufficiently good reason to do so. This could not be so unless the person presently has the bare ability to inflict harm, the ability to act on a will to hurt someone. If the person really were unable to hurt anyone, if he could not do it no matter how much he wanted to, then his present peaceable behavior would be involuntary. The behavior might be welcome, but it would not be to the person's credit unless he had somehow cultivated this inability to harm (assuming that this is a good thing). As I argued above, the ability to translate one's will into intentional action suffices for the ability to act on a will to refrain from that course of action.

The sort of person Wolf describes not only has the ability to do otherwise but has control over the exercise of that ability. He absolutely will not hurt anyone without sufficient reason to do so. "Couldn't do otherwise" commonly serves here just to emphasize the person's resolute moral judgment and control over his behavior. The remark means that, as things are and are likely to be, this person can be counted on not to do otherwise.

Perhaps most importantly, Wolf never goes very far in explaining why the notion she is examining is the notion of freedom. At the outset, she says that a free agent is one "whose actions are under his own control," "up to him to decide" (151). However, the freedom she is interested in cannot consist just in control, for the person suffering from a deprived childhood is less than fully free, on Wolf's view, though "this agent seems to have as much control over his life as we are apt to have over ours" (160). Why does this person's moral blindness upset his freedom? The only explanation Wolf gives is that this person is not blameworthy for doing what he cannot be expected to know any better than to do. Apparently Wolf holds that freedom is sufficient for being either blameworthy or praiseworthy, and this is why not being praiseworthy or blameworthy demonstrates unfreedom. But why is this? The connection is usually thought to run in the other direction: desert of praise or blame is sufficient for freedom. What is conspicuously missing from Wolf's essay is any discussion of why specifically *freedom* involves the competent appreciation of value. Maybe Wolf is really only examining praise- and blameworthiness. In that case, it is neither particularly illuminating nor

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

controversial to state that agents do not deserve praise or blame if they cannot understand the relevant values. It is generally accepted that we are to be excused for ignorance that we cannot reasonably be expected not to have.

Despite Watson's and Wolf's attempts, we remain in need of an adequate account of how—and why—values figure in free action.

III

We can gain a clearer sense of the need to supplement the control component of free action if we reflect on why the power of control is so important to us. One reason is this. The values by which our actions—and we ourselves—are assessed typically matter very much to us. This is due in large measure to our deep-seated desire to be able to justify our conduct.[11] But, if the norms governing our actions matter to us, then so will the ability to make our evaluative judgments effective in our actions be important. This ability, at a minimum, requires the power of control. Or, if Watson's view, understood as an account of control, is basically correct, this ability constitutes control. It is natural that the importance of our ability to regulate our conduct by means of our values would depend primarily on the importance we attach to doing what those values dictate we should do.

This explanation of the importance of control indicates why freedom might be thought to involve more than just control and what more it involves. For, if we care deeply about the value of our actions, then we want more than the power to translate our own value judgments into effectual willing. We also want to be able to appreciate the relevant values and arrive at competent appraisals of the alternative courses of action we face. Our concern for those values would be practically impotent if we could not bring them competently to bear in our deliberations about what to do.

We are now in a position to propose a condition of fully free action to supplement whatever condition accounts for control: one's action is fully free only to the extent that one has the ability to appreciate the normative standards governing one's conduct and to make competent critical evaluations, in light of those norms, of open courses of action. I will defend this condition by showing its connection with the still vague but enduring belief that a completely free act is fully our own. In fact, although this intuition can be used to show the appeal of the proposed condition, I will use the condition, in turn, to clarify

[11] Contractualist accounts of morality rightly give this desire a prominent role in moral motivation. For instance, see T. M. Scanlon, "Contractualism and Utilitarianism," in Amartya Sen and Bernard Williams, eds., *Utilitarianism and Beyond* (New York: Cambridge, 1982), pp. 103–128.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the intuition's content. The result should be a better grasp of the notion of freedom at issue and, furthermore, an argument for an unorthodox condition of full freedom. Alignment of the interpretation of the intuition with the widely accepted necessity of control for freedom will also speak in favor of that interpretation and lend more weight to the proposed condition.[12]

To begin, consider an ordinary enough example. In the course of a game of make-believe, a young child climbs up on my desk when I am not around and begins playing on it. She has not asked me whether she may play on my desk. She knows that I would want her to ask, because she knows that I do not like her to play there. She has played this game before, and I have warned her not to play it again. Of course, this history figures in the appeal the game has for the child on this occasion. The game is a little risky, so all the more entertaining. But, though the child deliberately plays on my desk without asking me, knowing that I really do not like her to do this, she is not mature enough to understand why I do not like her up on my desk or why this dislike, connected with the importance of a private work space, is reasonable. The girl is not just flailing around aimlessly in my office, lost in the "fog" of infancy. She is deliberately playing a game I prohibit. She has no trouble thinking to herself, "I sort of like being bad." Her childishness reveals itself in her inability to appreciate what is serious about my requests, about being "good" or "bad."

Does the child freely play on my desk? Before we answer, we might pause to wonder what there is about the case which should lead us to question the freedom of the child's action. It could seem that the question injects the issue of freedom into the circumstances so artificially as not to be a genuine question at all. And, if it is not really a proper question, then we need not bother trying to answer it.[13]

Even if it were odd to raise the question of freedom in the imagined setting, that would not mean that the question is not worth investigating. In any case, the question of the girl's freedom is not unnatural here. What attitudes I hold toward her (once I find out

[12] If control should turn out to consist in the ability to do what one thinks best, all things considered, then the proposed condition of normative competence, though logically independent of control, would complement the control condition in a natural way. Thus, the plausibility of one condition would significantly enhance the case for the other. A complete defense of each of these conditions would make use of this point. As the present essay concentrates only on the normative-competence condition, however, the defense offered for it here is unavoidably less complete than it would be in the context of a full theory of free action.

[13] This is the kind of stance a Wittgensteinian might take.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

what she is doing) may well depend, in part, on what I believe about her freedom on this occasion. And whether or not these attitudes are appropriate can depend partly on the truth of my belief about her freedom. If I am not merely upset with the child and inclined to scold her, but blame her for what she has done and am inclined to punish her for playing the game, then I must believe that the child acted freely. If I am to blame her justifiably and punish her fairly, then my belief about her freedom must be true. In making these claims, I am not supposing that adequate conditions of blameworthiness and justifiable punishment are immediately obvious. My present aim is to enlarge the scope of the traditional problem of free action. Few traditional theories have challenged the necessity of freedom for desert of blame and for justifiable punishment.[14] So, if there are certain attitudes that are standardly characterized in terms of their ties to free action, and the adoption of some of these attitudes may naturally become an issue in the circumstances of our example, then it may be quite appropriate to question the child's freedom.

A different reason for worrying that the question is artificial may be that the child's play seems so obviously free that it would be purely academic, in the most annoying way, to investigate its freedom. It is not because there is nothing to be said about the child's freedom or unfreedom, but because there is nothing very interesting to say about it, that the question may seem contrived. Everything said in describing the child's behavior suggests that it was entirely up to her whether or not she played her game. For all we know about her, she was neither acting on irresistible impulses nor so absorbed in fantasy as to lose control over whether or not she played on my desk. If anyone is ever free to do otherwise, this child seems to be. Indeed, the delight she takes in not heeding my warnings signals her own awareness of her power of relatively unimpeded choice.

I agree that, as far as we can tell, the child is in control of her action, that she is free to do otherwise. It does not follow that she is therefore obviously acting freely when she plays the game on my desk. This will be so only if control exhausts freedom. It is true that most writers have assumed as a matter of course that freedom consists in nothing more than control. We have seen that even Watson and (at least occasionally) Wolf believe this. However, although we

[14] Even the most crudely consequentialist theories of punishment traditionally have interpreted freedom so that the necessity of freedom for justifiable punishment is preserved. See Moritz Schlick's classic paper, "When Is a Man Responsible?," in Bernard Berofsky, ed., *Free Will and Determinism* (New York: Harper & Row, 1966), pp. 54–63.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

can grant that the child acts freely in this instance to the extent that she has control over what she does, if we admit that free action is a matter of degree, that our actions can be more or less our own in the pertinent sense, then it is open to us to contend that the child's action is not completely free.

This seems to me to be what we will say, on reflection, about the case. In the sense in which we feel that an unqualifiedly free action is fully the agent's own, the child's action is not fully her own. This is because she does not understand well enough the evaluative claims that attach to her playing on my desk. Believing that the child's freedom is diminished by her lack of competence in evaluating her own conduct is a logical companion for the widespread belief that a deprived childhood reduces freedom. Wolf has shown that we do not need to think that a deprived childhood somehow determines a person's future conduct as an ordinary upbringing does not, in order to sustain the belief that an abnormal upbringing can directly restrict freedom. We need only see that what deprived childhoods are deprived of, in this context, are those things which foster the development of the ability to evaluate our own conduct competently.[15]

The reduced freedom of agents like this child has not been given very close attention in contemporary discussions. Yet it has not been ignored entirely. Frankfurt, for instance, admits the unfreedom of "all very young children" and tries to show how this coheres with his thesis that the possession of second-order volitions is essential both to personhood and to freedom (of action and of the will) (11). Frankfurt maintains that all very young children are "wantons," agents who have no higher-order volitions with respect to their first-order desires. This explains, on Frankfurt's theory, why such children cannot act of their own free will.

Frankfurt's view confirms my sense that mere control is insufficient for complete freedom, young children being a prime illustration of this point. His explanation of it will not do, however. Although he is no doubt correct that we all start out as wantons, the particular case I am considering raises the possibility that our freedom may be incomplete even after we have begun regarding our first-order motives critically. The child playing on my desk may be old enough to begin forming values of her own, however unsophisticated. She may have a crude system of higher-order ends in terms of which she reflects on her ground-level wants. She knows that I some-

[15] See also Irving Thalberg, "Hierarchical Analyses of Unfree Action," *Canadian Journal of Philosophy*, VIII, 2 (June 1978): 211–226, esp. pp. 218, 221/2. Thalberg grants that a "brutal or deprived upbringing" lessens freedom, though he is unsure how to explain this.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

times disapprove of what she does. I see no reason why it would be impossible for her to begin forming similar attitudes toward her own actions and motives. She could do this before being able to comprehend the values on which my attitudes toward her rest. It seems highly counterintuitive to think that, before a child can move out of thoroughgoing wantonness and begin to develop her own second-order volitions, she has to appreciate adequately the much more sophisticated values of the adults around her. The view that the child, who is no longer a wanton but who cannot comprehend adequately the normative standards that apply to her behavior, is less than a fully free agent seems continuous with the view that no *very* young children can act freely. Frankfurt misses the deeper element of freedom which is at stake in these cases, identifying only one necessary feature of it, the capacity to be critically aware of one's own will (11/2).[16]

Considering the case of the child at play, Wolf's deprived-childhood example, and Frankfurt's wanton children, we become aware of an entire range of hindrances to complete freedom which traditional theories have ignored. Toward one extreme we find persons whose freedom is diminished because mental illness or disability, or sheer lack of needed education or care, makes them ignorant of the standards that apply to their conduct. Toward the middle of the range are persons like the child at play, who are aware of some relevant norms and their ordinary application but fail to appreciate why those values legitimately apply to them. This can also face normal adults when they find themselves in some special setting where they are unfamiliar with what they are expected to do. Nearer the other extreme are cases in which persons adequately appreciate pertinent values but are barred by special circumstances from bringing this to bear in their deliberations. Here freedom is reduced because agents cannot bring their normative insight into play in making the practical judgments that their power of control enables them to act on. This is one way in which the pressure of a dire emergency can restrict our freedom. All the cases in this range seem to fall under the proposed condition of freedom.

IV

We should return for a moment to one of the weaknesses I noted above in Wolf's argument. Wolf purports to be providing us with an account of why the power of free action depends on our powers of moral insight. She supplies us with examples, like the victim of the

[16] The ability to form higher-order volitions is necessary for the proposed condition of freedom, because in evaluating courses of action one may have to evaluate motives as well as behavior and its consequences.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

deprived childhood, in which praiseworthiness and blameworthiness are undercut by incapacity to evaluate courses of action competently. But why should we think that these examples show anything about freedom? Should we think that freedom is sufficient for blameworthiness, or for accountability in general? These questions also seem to threaten my claims about the necessity of a certain critical normative competence for freedom.

I hope to have advanced the case for the normative content of freedom beyond Wolf's argument by linking free action to the metaphorical—but nonetheless steadfast—intuition that such action is distinguished by being fully one's own, and then using that intuition to force theoretical treatments of freedom to address evaluative appreciation as well as control. Yet this will not dispel the fear that in the process we have strayed past the boundaries of our subject if we cannot say any more clearly what that distinctive sense is in which an unqualifiedly free act is fully the agent's own. Otherwise the alleged intuition will be suspect, perhaps only a thin disguise for our failure to comprehend the concept of freedom or for the incoherence of the subject itself.

It is not difficult to collect clues about what fully owning our actions does *not* involve. We have seen that it need not directly involve constraints on the causal determination of our actions. The question before us concerns the abilities we need to perform actions that fully belong to us. It may turn out that having these abilities entails that our actions are not causally explainable in some circumstances. I doubt that this is so.[17] In any case, this would make restrictions on causal explanation per se only indirectly necessary for full ownership of our acts.

It is also fairly clear that approving of our acts or judging them the best that we can do in the circumstances, all things considered, is neither necessary nor sufficient for their being fully our own. It is not necessary because acts that are fully ours may be ones we condemn. Unless weakness of will is an illusion, we can sometimes freely do what we believe we should not, on balance, do. Thinking our actions best is not sufficient because such acts may nonetheless be compelled. I find no reason to disturb the orthodox belief that the sort of intimate ownership of our actions in which freedom consists requires control over them.

An initial defense for the belief could go as follows. A compelled action is not fully ours because, even when we actually approve of it, we would have performed it anyway had we not valued it. The com-

[17] See my "Ordinary Ability and Free Action," *op. cit.*, for extensive argument on this matter.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pulsion dictates that we have the will to perform the action no matter what sort of person we are. The uncontrollable source of the action disallows certain sorts of normative judgment about us from being made on the basis of the action's character. The compulsion blocks the permissibility of inferences from particular kinds of evaluation of the action to corresponding evaluations of us. The action does not fully belong to us because, in the terms of whatever values apply to it, it cannot by itself reveal what kind of person we are.[18] Again, the fact that we might actually happen to endorse a compulsive will does not alter the fact that lack of control restricts agent ownership in this way.

This point moves us closer to an interpretation of the crucial intuition about freedom. Actions are fully our own when they may serve to express what we are like as persons, when they are potential vehicles of self-disclosure. That is, our free acts are suitable grounds on which certain sorts of evaluative appraisals of us may be formed. Free actions are especially our own in that qualities such as their maliciousness or kindness, their ineptitude or grace, their rudeness or courtesy can properly give rise to consonant critical judgments about us (either by others or by ourselves).[19]

One worry about the proposed interpretation of the ownership intuition is that it may not single out specifically free actions. The

[18] The distinction between evaluations of persons and assessments of their conduct which this point presupposes is especially common in the context of moral judgment. Thomas Nagel expresses a closely related distinction in a familiar manner:

> Prior to reflection it is intuitively plausible that people cannot be morally assessed for what is not their fault, or for what is due to factors beyond their control. Such judgment is different from the evaluation of something as a good or bad thing, or state of affairs. The latter may be present in addition to moral judgment, but when we blame someone for his actions we are not merely saying it is bad that they happened, or bad that he exists: we are judging *him,* saying he is bad, which is different from his being a bad thing.

See "Moral Luck," in *Mortal Questions* (New York: Cambridge, 1979), p. 25.

[19] Since at first it can seem strange to understand talk of ownership in terms of some notion of self-expression, it is worth recalling that a variety of associations between possession and personal disclosure appear in areas of discourse unconnected with free agency. Dangerous as it can be, we often seek evidence among her belongings about what someone is like as a person. People sometimes refuse to buy a particular sort of car or book or article of clothing for the related reason that "it is not really *them.*" Here the language of self-revelation is applied to property in virtue of the latter's potential for disclosure. Elsewhere the language of property can mark self-revelation, as when we distinguish uncharacteristic or affected behavior from more usual or authentic behavior by saying things like, "That's really her laugh," or "Was that tone really his?" Of course, linguistic conventions such as these do not establish my claim about the connection between ownership and normative self-disclosure in the case of free action. They merely suggest a philosophical hypothesis that seems worth exploring for independent reasons.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sorts of action which characterize nervous, excited, or lethargic people, for instance, may properly anchor inferences that they are nervous, excited, or lethargic persons even when those actions are not fully controllable and thus quite unfree. But the inferred assessments are of a kind from which we normally can excuse ourselves. By contrast, we cannot legitimately escape or shake off with pleas for excuse the sorts of potential self-revelation or evaluation which full ownership of our actions makes appropriate. Our free actions permit judgments about us which we can acceptably dodge only by defending against them, by offering justification for the actions that prompted them. Without such justification, we have to accept them. Our free acts are fully our own insofar as they potentially afford appropriate bases for normative assessments of us in the face of which we have no excuse.

This further clarifies why uncontrollable actions are not fully our own, or self-expressive, in the pertinent sense. Such actions may permit some judgments about us as persons—for instance, that we sometimes do things over which we have little control—and these may support various criticisms of us. But our being psychologically compelled to do something malicious, for example, would undercut any potential judgment that we are malicious persons, along with the moral censure that such a judgment might effect or embody. Because of its uncontrollability, our action would not support any moral disclosure of us. It would, therefore, be less than fully our own, not fully free, even if we really were malicious. The meanness of our action would only accidentally correspond to our actual meanness.

Similar reasons establish that the range of cases outlined above, in which agents' ability to evaluate their conduct may be underdeveloped or diminished, displays another dimension along which freedom can be limited. Where agents cannot capably evaluate their conduct, that conduct is not an appropriate basis for making the relevant sorts of assessment of them as persons. Since the child playing on my desk cannot be expected to appreciate why it is wrong to do so without asking me, appropriate evaluations of her conduct cannot properly ground corresponding evaluations of her. She does something that is inconsiderate, disrespectful, unkind. Yet these qualities do not directly express or disclose what sort of person she is; for (from what has been said about her) she is neither inconsiderate nor disrespectful nor unkind. Her action cannot reveal what she is in these terms, because, although her action is inconsiderate and unkind by applicable standards, those standards do not apply to *her*. As the child cannot presently understand what consideration means and why it is valuable, we would get only a distorted picture of her from

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the ways in which her actions intrude on others' interests. This is why the child's playing is less than fully her own. Relevant normative criticisms of her conduct cannot reveal who she is in such a way as to permit corresponding criticisms of her, criticisms in the face of which she has no excuse.

It by no means follows that the child's present personal qualities have no explanatory ties to her unwelcome conduct. One reason why the child plays her game is probably that she does not appreciate how it matters to me. In a weak sense, then, the inconsiderateness of the game may express what the child is, for the deficient state of the child's evaluative ability may be an explanation for the shortcomings of her conduct. The stronger sort of self-disclosure which distinguishes fully free action requires that normative judgments of the action can correctly support inferences to evaluations of the person which employ the same terms as the former judgments. An action does not qualify as free merely by showing something or other about its agent. It is compatible with less than fully free action that the action reveal significant features of the agent's will and values. The child's game shows that she wants to play more than she wants to avoid my scolding. It shows that she values the quality of her play more than she values whatever the things on my desk mean to me. These conclusions are possible because of the child's control over her action. And, as we noted above, it is trivially true that even uncontrollable actions can tell us important things about their agents. But our actions cannot be fully our own in the sense freedom demands unless normative properties we possess can be disclosed in corresponding normative properties of our actions. Critical evaluative ability is a condition of freedom in virtue of this dimension of full ownership of action.

This should answer the objection I put before Wolf. Expanding a theory of free action to include a substantive evaluative ability as well as control does not disguise a departure from the original subject. These conditions belong together because both are necessary for fully owning our actions in the specified sense.

Two small clarifications, and the basic case for the proposed condition of freedom is complete (though see fn 12). First, this condition should not be restricted to certain kinds or domains of value. The question of our ownership of our actions can just as reasonably arise against the background of standards like the rules of games or rules for the correct use of a language as against the background of moral standards. Hence, the values that a free agent must appreciate and be able to apply in her deliberations can include values other than moral values. This allows, however, that, if the sort of evaluation of persons we normally most care about is moral evaluation, then moral

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

understanding would be most important for having the freedom we most care about.

The second point is that the proposed condition does not entail that any free action must actually be a reasonable object of critical appraisal at, or shortly before, the time of its performance. Many of our free acts—drumming our fingers, tying our shoes, ordering coffee instead of tea—are ordinarily so insignificant and plainly acceptable, by whatever standards of conduct apply, that it would be silly to reflect critically on the value of performing them. This is why the proposed condition should be taken to mean only that free acts are potentially reasonable objects of critical evaluation. (For the same reason, I have avoided saying that free actions are always reasonable grounds for making the relevant sorts of assessments of their agents or are always actual vehicles of the relevant sorts of self-revelation.) If it happened that there were some important reason why we should not drum our fingers, then there might be sufficient grounds for reflectively evaluating that action. But, in that case, it would not be the freedom of the action that changed, only the significance of its freedom.[20]

V

I shall conclude by responding to some likely objections.

One could object that, because the normative-competence condition largely concerns knowledge of applicable norms, it might be merely a special case of the intentionality demanded by control.[21] Having control over whether or not one performs an action demands that one know what one is doing. But, according to this objection, when one is not aware of the norms that apply to one's conduct or does not possess the concepts necessary to understand them, one will not really know what one is doing. One's freedom will be undermined by the unintentionality of one's action (appropriately described), not in virtue of any independent normative condition. The girl playing on my desk is unfree, on this view, because she does not grasp the concept of respecting important work and so damages important work unintentionally.

I accept that some of the things we do may be unintentional if we suffer from underdeveloped or diminished evaluative ability. The objection fails, however, if we can produce cases in which actions fully under their agents' control are less than fully free because of

[20] I am grateful to the editors for prompting me to strengthen the argument of the preceding section.

[21] This objection was suggested to me by Alan Goldman, who commented on the earlier version of this paper delivered at the Central Division Meeting of the APA in May, 1986.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the agents' failure to meet the normative-competence condition. I have produced such a case already. The child may damage important work of mine unintentionally. This is not the action I was specifically concerned with, however. Whether or not she damaged important work (and she might not have), she played on my desk without asking me; she played a game I prohibited. Under these descriptions, the child acted intentionally and was in control of what she was doing. What restricted the girl's freedom was her inability to appreciate the reasons why the game that she knew I disapproved of really was wrong. The required normative competence does not collapse into intentionality.

A second objection is directed at my interpretation of the full ownership of action which intuitively characterizes free action. The objection is that, by associating such full ownership so closely with criticizability for one's conduct, the proposed account dissolves freedom into accountability or responsibility. Many find it implausible that freedom should be sufficient for accountability and absurd that these should be one and the same thing. My argument might be thought to support both of these objectionable claims.

I have already tried to dispel much of the force of this objection. Its appeal seems to derive mostly from the dogmatic conviction that freedom simply cannot consist in anything more than control or ability to do otherwise. The mere mention of evaluative insight triggers the reaction that the topic is really responsibility. My main response has been to link both control and evaluative competence to a shared intuitive background that itself can be understood in a way that reflects why freedom is so important to us. Still, the objection that I have blurred freedom and responsibility is so tenacious that I might say something more in response.[22]

We should attend carefully to the distinction between being potentially open to critical appraisal on the grounds of one's conduct and being a reasonable target of such criticism. I noted earlier that free actions are often not reasonable bases for evaluating their agents; free acts are only *potential* vehicles of the relevant kinds of self-revelation. When we are responsible or accountable, however, there must be something that we can be held responsible or accountable for. There must be a charge for us to answer or something favorable to count to our credit. In general, there must be some reason for us to be appraised because of what we have done, though there need not be sufficient reason to blame or praise, condemn or admire us.

[22] The deep appeal of this objection first became evident to me in discussions I had with T. M. Scanlon.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

This is not so, on my view, for merely free acts. These are often so insignificant as to give no reason whatsoever for us to be charged or credited for performing them. Acting freely need not, and commonly does not, give us anything to be responsible or accountable *for*. Admittedly, the interpretation of freedom as an evaluative notion directs our attention to some features of agency which have normally figured in theories of responsibility. But, if theories of free action should begin to draw much more heavily from value theory, as I contend, then trespass of the traditional boundaries between theories of freedom and theories of responsibility is entirely warranted.

A final objection is that the application of the proposed condition of freedom has been left ill defined. In any particular situation, the condition carries specific normative content: the agent must appreciate applicable norms and be able to evaluate alternative courses of action in terms of those norms. But where many distinct sets of standards govern our conduct, as is typically the case, which standards should matter for determining our freedom? Is our freedom less than complete if we are not adequately cognizant of *all* these standards? This is the answer the foregoing discussion suggests, though it may not be satisfactory. If I meet the condition with respect to moral standards, but fail to meet it fully in respect of some other domain, such as etiquette (my childhood may have been deprived of those things needed for the cultivation of a comprehension of correct table manners), then why should my freedom as a moral agent be affected? It seems strange to allow my blindness to what makes for a well-mannered dinner guest, however excusable, to erode my freedom on all those occasions where good table manners are called for.

I do not pretend to have a complete resolution of this issue. I think we should consider the following possibility, however. The normative competence condition rather naturally leads us to believe that the freedom of our actions is most perspicaciously addressed from the standpoint of some specifiable normative domain or other. Confusion about how to apply the condition can be dispelled, then, if we let go of the hallowed idea that the freedom of our conduct is a factual matter independent of the well founded, normative standpoints from which we might be moved to probe into its freedom. There may be no such thing as the freedom of an action *simpliciter*. There may only be freedom in relation to specifiable domains of value. If this is so, then the grammar of freedom is relational.

PAUL BENSON

University of Dayton

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

A FIRST LOOK AT THE PORNOGRAPHY/CIVIL RIGHTS ORDINANCE: COULD PORNOGRAPHY BE THE SUBORDINATION OF WOMEN?

FOR a philosopher immersed in the analytic tradition, the "pornography issue" may be puzzling. This issue has been subjected, as the years of its tenure *as* an issue have increased, to a more and more complex analysis. The very definition of pornography has expanded, almost geometrically, from a few words to a few paragraphs to a few pages.[1] Early in the literature—a literature which now seems as quaint as the cosmological visions of the pre-Socratics—the general motivation for discussion and analysis seemed to be to explain why "dirty pictures," though perhaps not everyone's cup of tea, were not particularly harmful,[2] and were certainly nothing that a sane, liberal-minded society would restrict *by law*[3] (Lord Devlin's model society not being sane in its reliance on subjectivism and certainly not liberal-minded in its legislation of the good[4]).

In those dear dead days of simplicity, to have suggested that pornography itself (and not its censorship or actionability) had a direct relation not to questions of mere value or virtue but to questions of *justice*—and justice in the strictest deontological sense, not some utilitarian simulacrum—would have been seen as confused, irrational, or duplicitous.[5]

[1] An example of a fairly standard (medium-length, too) definition of pornography would be Longino's: "Pornography . . . is verbal or pictorial material which represents or describes sexual behavior that is degrading or abusive to one or more of the participants *in such a way as to endorse the degradation.*" Helen Longino, "What Is Pornography," in Laura Lederer, ed., *Take Back the Night* (New York: William Morrow, 1980), p. 43. Some problems with this idea of "endorsement"—problems tangential to my inquiry—are noted by Alan Soble in "Pornography: Defamation and the Endorsement of Degradation," *Social Theory and Practice*, XI, 1 (Spring 1985): 61–87.

[2] And of course some claim these representations are socially and personally helpful, sexually liberating, and so on. See, for example, G. L. Simon's, "Is Pornography Beneficial?" in Thomas A. Mappes and Jane S. Zembaty, eds., *Social Ethics* (New York: McGraw Hill, 1977), pp. 243–248.

[3] This is not to say that otherwise sane and liberal–minded societies have not restricted pornography by law, for of course they have. The arguments of liberal theorists—see, for example, Ronald Dworkin's *Taking Rights Seriously* (Cambridge, Mass.: Harvard, 1977), especially ch. 10, "Liberty and Moralism"—stress that restrictions against such material, under the aegis of legislating the good, are inconsistent with liberal political theory.

[4] See his *The Enforcement of Morals* (New York: Oxford, 1965), ch. 1.

[5] Again see Dworkin, especially ch. 11, "Liberty and Liberalism," p. 262, on which page Dworkin states that showing the movie *Deep Throat* is not a threat to any principle of justice.

0022-362X/87/8409/0487$02.50 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Then feminists began looking at pornography, most of them writing outside the analytic tradition and many writing outside of any (academic) philosophical tradition whatever.[6] Articles were written, definitions abounded, these were criticized and (usually) expanded; pornography was examined from every (as it were) angle. On some accounts, pornography started to look pretty bad.[7] (Of course, in the intervening years, pornography itself had got worse, more violent, more "sick," and so on.) Still, the question of what all this—'this' representing perhaps a state of extreme disvalue—had to do with *justice* was not made clear.

THE ANALYSIS OF PORNOGRAPHY: A HISTORY THAT ENDS IN A CATEGORY MISTAKE?

Now we have, as I see it, the conceptual culmination of this newer feminist investigation into pornography encoded in what is most often referred to as the Dworkin/MacKinnon ordinance.[8] The ordi-

[6] See, for example, Susan Griffin's fairly high-flying *Pornography and Silence* (New York: Harper and Row, 1982).

[7] The by-now classic investigation and analysis of pornography is Andrea Dworkin's *Pornography: Men Possessing Women* (New York: Perigee, 1981). See also the chapter "Anti-Feminism," in her *Right-wing Women* (New York: Perigee, 1982). *Take Back the Night, op. cit.*, is an anthology of women, all self-identified as feminists, writing about and criticizing pornography.

[8] The title "Pornography/Civil Rights Ordinance" is preferred by the ordinance's supporters, whereas the title "Dworkin/MacKinnon Ordinance" is preferred by its detractors. I use the two titles interchangeably.

This ordinance implies that there is a connection between pornography and justice, in that producing and selling pornography (as well as other pornographically related activities) is seen as a violation of the rights of (usually) women. I will touch only tangentially upon this question of justice in this essay, since I believe other questions having to do with the possible conceptual status of pornographic representations come first.

The ordinance defines pornography as follows (This is taken from the version—the Ordinance has several versions—reprinted in *Ms. Magazine* (April 1985): 46):

1. Pornography is the graphic sexually explicit subordination of women through pictures and/or words that also includes one or more of the following: (i) women are presented dehumanized as sexual objects, things, or commodities; or (ii) women are presented as sexual objects who enjoy pain or humiliation; or (iii) women are presented as sexual objects who experience sexual pleasure in being raped; or (iv) women are presented as sexual objects tied up or cut up or mutilated or bruised or physically hurt; or (v) women are presented in postures or positions of sexual submission, servility, or display; or (vi) women's body parts—including but not limited to vaginas, breasts, or buttocks—are exhibited such that women are reduced to those parts; or (vii) women are presented as whores by nature; or (viii) women are presented being penetrated by objects or animals; or (ix) women are presented in scenarios of degradation, injury, torture, shown as filthy or inferior, bleeding, bruised or hurt in a context that makes these conditions sexual.
2. The use of men, children, or transsexuals in the place of women in (1) above is pornography for the purposes of this law.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

nance was written by Andrea Dworkin, a feminist writer, and Catharine MacKinnon, a feminist lawyer. Dworkin and MacKinnon had been hired by the city of Minneapolis to develop an ordinance that would help the city control the problems related to "adult" material. The ordinance they wrote was passed by the city council, but vetoed by the mayor. It has since gone through various court battles at various levels of the American legal system, being most recently summarily dismissed by the Supreme Court. No doubt some form of it will surface again. In any case, it is not the ordinance's legal status that interests me here. I am interested rather in its status as containing a possibly meaningful philosophical claim. "Pornography," the ordinance states, "is the graphic sexually explicit subordination of women. . . ." This definition of pornography should make any English speaker sit bolt upright. Note the use of the 'is' of identity between 'pornography' and '. . . subordination'. Now that is peculiar. And the peculiarity is not, as is usually the case in law, a function of mere legal jargonizing—the party-of-the-first-part sort of talk—but is *conceptual*.[9] And because the peculiarity is conceptual, it is, I think, philosophically interesting.

A befuddled reader might sensibly react to the Dworkin/MacKinnon definition of pornography as she would, say, to the claim, "Green is where the post office can be found." This statement (or "statement") is also conceptually "off," but not *horribly* so. In fumbling with its near-sense, the mind, being the mind, tries to revive the claim, as in, "Well, the remark must here refer to a board game, like Monopoly, and it is upon the green square that the game's post office is situated." Now the mind can rest, having made complete sense of the remark. In the Dworkin/MacKinnon case, the reader striving for sense might translate the given definition by mentally adding a few words, as in "Pornography is the graphic sexually explicit *depiction of the* subordination of women." Now the definition makes perfect sense. But this amended definition is not what the ordinance provides. The ordinance says that pornography is the subordination of

[9] As evidence of the ordinance's conceptual difficulty, one can look to the sympathetic, but adverse, opinion of Judge Frank Easterbrook [quoted in *Off Our Backs* (April 1986): 6]: "Depictions of subordination tend to perpetuate subordination. The subordinate status of women in turn leads to affront and lower pay at work, insult and injury at home, battery and rape on the streets . . . (but) this simply demonstrates the power of pornography as speech." Clearly, Easterbrook mentally translated the ordinance's definition of pornography into a more sensible form, ignoring the copula of identity between 'pornography' and 'subordination'. The ordinance does not merely say that, in depicting subordination, pornography causes subordination. The ordinance says that pornography *is* subordination.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

women. Has the analysis of pornography ended, thanks to these feminists, with something like what Gilbert Ryle[10] called a category mistake or, to use A. C. Baier's term, a "semi-sentence"?[11] Surely, according to what I call the "reasonable view" of pornography, it would seem so.

PORNOGRAPHY: THE REASONABLE VIEW

The reasonable view of pornography remains straight and strong, the bones beneath the pile of paper flesh generated on this issue. The reasonable view has an incredibly deep and tenacious hold on our common sense and its near relative, our common moral consciousness. This may be because the reasonable view is true or because, though it is not true, it is conceptually welded to something else which is, and we cannot see the point of separation between this coin's true head and its false tail.

Like all moral views, the reasonable view is not unrelated to an epistemology and a metaphysic,[12] though these are not (as they usually are not) directly stated. The reasonable view is simply this: "Pornography is sexually explicit material—pictures on paper or film, or words on a page. Some people enjoy looking at this material; others do not. To forbid by law the production and dissemination of this material[13] is an act of moralistic piety gone unconscionably governmental—or outright tyranny. If you don't like the words, citizen, just don't read the words. If you don't like the pictures, friend, just don't look at the pictures." With some minor adjustments for principles of public offense[14] and property devaluation,[15] pornography on the reasonable view is rather quickly found to be something outside

[10] *The Concept of Mind* (New York: Barnes & Noble, 1949).

[11] "To make a new point it may be necessary for the philosopher, as much as for the poet or the scientist, to speak in a new form of words not simply translatable into any of the old forms. But the philosopher who speaks too often in semisentences runs the risk of semiunderstanding from only a semiaudience." See "Nonsense," in Paul Edwards, ed., *Encyclopedia of Philosophy* (New York: MacMillan, 1967).

[12] In both "Feminism, Marxism, Method, and the State," in *The Signs Reader* (Chicago: University Press, 1983), pp. 227–256, and "Not a Moral Issue," *Yale Law and Policy Review*, II, 321 (1984): 321–345, Catharine MacKinnon maintains that objectivity, in the sense of aperspectivity, is the way in which gender males both create and describe the world. From the point of view of such "aperspectivity" only direct, John-hits-Mary causal harm is recognized as the harm that creates injustice. Thus, if no such harm is connected with producing pornography, such production cannot be unjust.

[13] Or—which is different—to make it actionable in the manner of the Pornography/Civil Rights Ordinance.

[14] See Joel Feinberg's "The Offense Principle," in Mappes and Zembaty, *op. cit.*, pp. 252–257.

[15] See the case of *Young v. American Mini Theaters, Inc.* described in Archibald Cox's *Freedom of Expression* (Cambridge, Mass.: Harvard, 1981), p. 34.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the area of deontological note.[16] The *ease* with which this conclusion is reached might itself give us pause. We might reflect on this ease and we might recall one of Ludwig Wittgenstein's many suggestive remarks: "The decisive moment in the conjuring trick has been made, and it was the very one that we thought quite innocent."[17] So, on the one hand, we have the simple and reasonable view of pornography and, on the other hand, what looks to be a conceptually bizarre view, the Dworkin/MacKinnon view—something coming, almost, from another form of life. And here we may recall another dark saying of Wittgenstein's: "If a lion spoke, we could not understand him" (§7). If Dworkin/MacKinnon is any indication of the radical feminist form of life, some might be tempted to ascribe the same leonine inscrutability to its opaque practitioners.

THE LAW OF NONCONTRADICTION APPLIED TO OUR CASE

Now I assume that, unless we adopt some kind of fashionable relativism,[18] if one of these views of pornography is correct—the reasonable view or the Dworkin/MacKinnon view—the other is not. Either pornography is, at worst, a bad thing[19] which we would rather not have about but which justice requires us to tolerate, or it is, in addition to being a bad thing, something whose production or dissemination is an actual injustice of some particular sort in that it violates the rights of (usually) women.

Although I am, like, I imagine, almost everyone else, strongly attracted to the reasonable view (because it *is* reasonable), the bizarre view expressed by Dworkin/MacKinnon is to me of greater philosophical interest, because and not in spite of its conceptual peculiarity. This peculiarity suggests an entirely new paradigm, a virtually different world—or extreme mental confusion. Leaving open the possibility that it might be the former, I have begun an investigation. Naturally, one cannot examine this "different world"

[16] Again, I think that obscenity law, in legislating matters related to the good, is a deontological aberration. The just act/good act distinction is severely sketched by Charles Fried in *Right and Wrong* (Cambridge, Mass.: Harvard, 1979): "But while the demands of justice are implacable within their proper domain, it is . . . inappropriate and unnecessary to extend them outside of that domain. [There] the scale of judgment is marked, if at all, in degrees of praise only" (173 and 176). In other words, as long as what we do is within the bounds of just action, it is no one else's business adversely to judge, much less to prohibit or restrict, what we do. The production of pornography—as pornography is ordinarily defined by obscenity laws—falls into this realm of personal discretion.

[17] *Philosophical Investigations,* G. E. M. Anscombe, ed. (New York: MacMillan, 3rd ed., 1958), §308.

[18] A sensible non-Platonic defense of normative realism is given by Thomas Nagel in *The View from Nowhere* (New York: Oxford, 1986), especially in the section "Realism and Objectivity," pp. 138–143.

[19] Again, some claim it is a positively good thing.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and its plausibility as a conceptual map all at once. In what follows below, I begin, in a typical, analytic, immoderately small-potatoes way, to begin to try to determine the answer to what seems to me to be a precedential question, viz., Under what conditions, if any, can a representation (here, a pornographic representation) *literally do*[20] what Dworkin/MacKinnon claims pornography does, that is, subordinate women? Whether or not these conditions of subordination do or could hold in this or any other possible world is, naturally, a subject for another time.

AN EQUIVALENCE

If, as Dworkin/MacKinnon claims, "Pornography is the . . . subordination of women," then I do believe it follows that pornography subordinates women. (If x is the subordination of y, then x subordinates y.) And the reverse entailment is also true; that is, if pornography subordinates women, it follows that pornography is the subordination of women. (If x subordinates y, then x is the subordination of y.) Thus, pornography subordinates women if and only if pornography is the subordination of women. (x subordinates y if and only if x is the subordination of y.) My reason for pointing out this equivalence is heuristic, since it seems that the "shallow" or ordinary grammatical structure of English makes the 'x subordinates y' formulation more linguistically ordinary, more idiomatic, than the 'x is the subordination of y' formulation.

AN UNWELCOME GUEST

Much to my dismay as an analytic lover of the circumscribed, I found that, in order even to wave at or gesture toward an explanation of either of the above equivalent claims, I was required to turn my attention to the examination of some seemingly unrelated and fairly ragged notions. So now I turn to what I was disgusted[21] to find climbing aboard my conceptual boat—viz., to the notion of a *practice*.

PRACTICES AND THEIR CONSTITUENTS

A practice is a socially established, socially orchestrated human activity that aims at certain goals or internal goods.[22] Ideologies[23] provide

[20] Of course, the notions of the literal and the metaphoric are problematic. I am working here with our intuitive and imperfect understanding of these concepts. Unless one wants to stack one's philosophical deck, it is important not suddenly to move the line between what counts as literally subordinating someone (versus metaphorically subordinating her) when examining the case of pornography. Cf. Nan Hunter's comment on the metaphoric nature of pornographic subordination in "Is One Woman's Sexuality Another Woman's Pornography?" *Ms.* (April 1985): 123.

[21] Disgust is, of course, the proper analytic response to this ragged world's failure to be amenable to the strengths of one's taxonomic tools.

[22] My notion of a practice draws from that of Alasdair MacIntyre [See his *After*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the conceptual support for practices, and institutions are the arenas in which practices are manifested and which materially support those practices. Although not every human activity is part of a practice, very many, perhaps most, are.[24] Practices may have more or less explicit sets of rules and procedures, be more or less generally recognized as practices, have a more or less direct tie to "natural" needs and desires, and cover a more or less broad area of human endeavor. Examples of practices are baseball, medicine, poetry, film-making, still photography, racism, militarism,[25] and sexuality.[26]

Within every practice, just as within the world at large, there are those things which occur (e.g., events and actions) and those things which exist (e.g., individual persons or objects, states of affairs). To say this is just to say that the world of the practice is categorized by the same ontological cookie cutter as the rest of the world—whatever that cookie cutter is. This is to be expected, since of course the practice *is* just part of that larger world. The difference, however, between, say, an event that is part of a practice (e.g., striking out in the practice of baseball) and an event that is not part of a practice (e.g., a tree falling on an uninhabited island) is that the former event

---

Virtue (Notre Dame, Ind.: University Press, 2d ed., 1984), p. 187], but MacIntyre seems to regard practices as functionally related to the human virtues, and I do not.

[23] My (fairly stipulative) notion of an ideology is such that it includes *all* of a practice's conceptual support, such as its rules, principles, value judgments, and purposes. These may or may not be known by the practitioners. Thus, nothing in my notion of an ideology requires that it be either covert or oppressive. This somewhat extended notion of an ideology allows us to see the structural similarity between social units like baseball and social units like racism. It also allows us to see how ideologies partially overlap to link practices together to form a single social reality —e.g., baseball and racism may share some conceptual support in their ideological inclusion of value judgments about the desirability of hierarchy or competition. A concise introduction to both the Marxist and the non-Marxist concepts of ideology can be found in David McLellan's *Ideology* (Minneapolis: Minnesota UP, 1986).

[24] I do not believe that this near ubiquity extends the concept of a practice into meaninglessness. It is not that the concept is either overbroad or empty, but rather that practices are as common as societies themselves are.

[25] Not everything that is a practice has a convenient, unambiguous English word to refer to it. 'Racism' and 'militarism' for example, might be taken by some to refer to ideologies, rather than practices.

[26] Michel Foucault's writings on sexuality make this point clearly. See his *The History of Sexuality* (New York: Vintage Books, 1980) and *The Use of Pleasure* (New York: Pantheon, 1985).

To say that an activity is a practice is not, of course, to deny that it incorporates certain "natural" needs and desires. Baseball, though a practice, certainly incorporates certain "natural" needs and desires (e.g., for physical exercise). It is in fact hard to imagine a practice that did not, in some way, incorporate human needs and desires. This incorporation is at least part of the reason the practice exists. If we keep this fact of incorporation in mind, the seeming peculiarity of calling sexuality a practice disappears.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

gains its identity, not as an event, but as the event it is—gains, that is, its meaning, significance, and characterization—in virtue of its being a part of and informed by its practice. All the constituents of a practice—its events, objects, actions, and so on—are what I call *practice informed.* That is, their identity—their meaning, significance, and characterization—is a function of their practice. In order to stress the practice-based nature of that identity, I refer to these constituents, generally, as *practice constituents* and, specifically, as *practice actions, practice objects,* and so on.[27]

The identity of a practice constituent can be traced backward to its generation by the practice's ideology and forward to a manifestation in the practice's supporting institution(s). For example, in the practice of medicine, the identity of the practice action of surgery—what surgery means within the practice, what counts as surgery, what characterizes surgery as surgery—is generated by an ideology that includes beliefs about the causes of disease, the proper and improper techniques of cure, the paradigm of human health, and so on, and the supporting institutions (e.g., hospitals, doctors' out-patient facilities) display or manifest this practice-identified surgery. Although the practice posits benign intentions on the part of the surgeon, as well as some degree of skill, that the individual surgeon may have neither benign intentions (he intends to make money by performing an unnecessary procedure) nor great skill (he's a bumbler) does not make his action any less surgery. The identity of his action is given by the practice, not by the individual practitioner. Even if he directly intends to kill you, he kills you *by performing surgery,* for that is the practice identity of his action.[28]

[27] To characterize this identity partially in terms of meaning and significance is already to indicate that I am applying a notion of identity which internalizes social context. (In which case, to call it "context" is rather misleading.) In finding the identity of practice constituents to be a function of their practices, I am denying the meaningfulness of a concept of identity free of social context, such as that said to characterize what have been called "basic actions." [For a discussion of the problems associated with basic actions see Michael Simon's "The Social Nature of Action" in his *Understanding Human Action* (Albany: SUNY Press, 1982 pp. 24–40).] I am further extending the concept of an identity which internalizes social context to objects and individuals as well. Though I do not do so here, this same application of a context-internalized identity may also be successfully extended to feelings and other mental states—see, for example, Robert Kraut's "Feelings in Context," this JOURNAL, LXXXIII, 11 (November 1986): 642–652; or Naomi Scheman's "Individualism and the Objects of Psychology," in Sandra Harding and Merrill Hintikka, eds., *Discovering Reality* (Boston: D. Reidel, 1983).

My account of a context-internalized identity differs from that of Simon and Kraut only in that I stress a smaller social unit, a practice, as being the generator of that identity, rather than society or social reality at large.

[28] A surgical tool is an example of a practice object whose identity, though certainly practice informed, has a connection to the practice's ideology which is both

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The practice-informed identity of practice constituents is nonsubjective and determinate. For example, in the practice of baseball, the third futile swing at a good pitch is the practice action of *striking out*. That the third futile swing at a good pitch is that action of striking out is not a matter of individual opinion. (Of course, it may be an open question whether or not the pitch was good.) It is also not a matter of opinion whether the batter who struck out has been bested by the pitcher or not. If he struck out, he has been bested by the pitcher. The practice identity of the action is such that one can say with certainty that the third futile swing at a good pitch is striking out, and that the unfortunate batter who performs this action is in fact bested by the pitcher. The batter on his third futile swing cannot change the identity of his action by closing his eyes and repeating to himself, "I am not striking out, I am not striking out," or by any other change in attitude or behavior.

The nonsubjective identity of practice objects follows this same pattern. In the practice of baseball, a certain type of wooden or metal object is a *bat*. That is the identity of that practice object. To claim, as some might be tempted to do, that the material substratum "beneath" a practice object (or action or event) is its *real* identity—to claim, for example, that a bat is *really* just a piece of wood—is a confusion and sometimes perhaps a lie. There is no more proper, basic, or real identity than practice identity. To provide an extreme example, someone who used an oven from Auschwitz as a punch bowl, claiming it was *really* only a waterproof container, would be someone caught up in this confusion about "real" versus practice identity. Logically speaking, there cannot be many identities for a

---

weaker and more indirect than the connection between that ideology and the practice action of surgery. Generally, the connection between an ideology and the identity of practice actions is both more direct and stronger, while the connection between ideology and the identity of practice objects is more indirect and weaker. Put in a linguistic mode, the move from some ideological statement or statements, e.g., 'The purpose of surgery is to cure disease' and 'This is a scalpel' is more roundabout and attenuated than the move from 'The purpose of surgery is to cure disease' and 'This action is surgery'.

The differing degrees of the tightness and strength of these ideological connections is, I believe, revealed in our intuitive, commonsense judgments that, for example, what the surgeon does can only be identified as surgery, while his tools, put to a different use in a different practice, might easily and properly be thought of as having metamorphosed into entirely different things (like garden clippers). Because of what is *in general* this stronger connection between a practice's ideology and practice actions, practice actions (as opposed to practice objects) are not good candidates for an attempted re-identification/metamorphosis through a new practice—for as long as the old practice and its ideology are extant (or even remembered), the metamorphosis to a new identity will be unlikely to occur.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

single practice constituent, and, given that we cannot opt out of human practices by fiat, practice identity *is* identity. A bat is, nonsubjectively, a bat, not "really" a piece of wood. And the qualities needed in order to be a good bat, or a good batter, are also and similarly nonsubjective practice-functions. In those cases in which it might be said that "the same" object or event is part of two conflicting practices—e.g., a raised arm might be a salute or a request for a teacher's recognition—the practices themselves will generally have procedures for establishing the legitimacy or precedence of their own informing power. (This stamp of the practice may or may not be governed by explicit rules.) At the mass political rally, the raised arm is a salute[29]; in the classroom it is a request for recognition.

It is possible for two conflicting practices to lay simultaneous claim to an event or object, but, in this conflict of meaning or identity, the victory of the one practice over the other is never complete. The cloud of the lost meaning threatens reidentification at every moment. Survival of an unwelcome practice might actually depend on this very instability and confusion. At the moment the malevolent identity of the practice constituent becomes nearly evident, the unwelcome practice can hide the actual identity of its constituent behind the more palatable identity generated by the conflicting practice. ("No, they are not saluting the Führer, they are merely asking for permission to speak.") And the wolf in sheep's clothing looks very like a sheep.

The indications of this dissimulation—of one practice hiding an identity behind another—are various. One clue is the effect of the use of "really" locutions, this effect being the production of an incoherence in a practice-informed explanation. ("If they are *really* just requesting recognition, then why is no one being called upon?") Another can be found in a break or gap in what is ordinarily a discernible, solid line connecting the practice's ideology and the alleged identity of the action or object, a break which may itself be recognized by the bizarre effects of the practice constituent (e.g., claims for teacher recognition should not induce mass chanting). Practices that rely on this type of dissimulation are properly suspect. We are led to wonder what practice identities they are hiding, and why.

A point related to the above is this: since the identity of a practice constituent is practice informed, it follows that this identity is a

[29] This may be so even if the agent does not intend to salute. "[T]he fact that an action must be subjectively or intersubjectively meaningful need not be taken to imply that its meaning has to be what the agent thinks it is" (Simon, p. 37).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

function of the given, extant practice, whatever it is, and not of some other ideal but nonextant practice, *even if that unrealized practice is objectively preferable.* For example, even if four-strikes-and-then-out baseball might be preferable (morally or rationally) to our present game, the third futile swing at a good pitch is still, nonsubjectively, the action of striking out. To say of a batter on his third futile swing, "He is not *really* striking out because it would be ever so much better if four strikes, not three, were out," is, at best, an extremely misleading way of saying that the present practice of baseball should be altered. And, similarly, to say of this unfortunate batter that he has not *really* been bested by the pitcher because baseball as we play it is a terrible game, or should be abolished, or only placates the workers, or whatever, is just false. The value predicates and characterizations that modify practice constituents are, just like the factual predicates and characterizations, functions of the given practice. A claimed reversion to an *ideal* practice identity for some constituent is the most common form of the type of dissimulation discussed above. But it is foolish and sometimes dangerous to forget that it is the real practice that makes the reality.

DEPICTIONS OF PRACTICE CONSTITUENTS

A depiction of a practice constituent, like the practice constituent itself, gains its nonsubjective identity from whatever practice or practices inform it. The fact that the various kinds of depiction making are *themselves* practices entails that the identity of a depiction may be *doubly* practice informed because it exists in an area, if not of *conflict,* then of *overlap* between two practices. The depiction-making practice may add something to the identity of the depicted constituent, or it may function like a blank tablet or mirror on which the depicted practice merely makes its identifying mark. A photo *of a baseball bat*[30] is a photo of a baseball bat because the practice of baseball informs the identity of that photograph. Because the practice of baseball informs the identity of that photograph, the picture is really a picture of a baseball bat, and not "really" a picture of a piece of wood, just as the bat itself is really a bat, and not "really" a piece of wood. Similarly, a film of a batter on his third futile swing is really a depiction of a batter striking out, and not "really" a depiction of certain body movements of a man in funny pants. Just as, given an extant nondepictionary practice, the practice constituents of that practice are nonsubjectively identifiable and characterizable, so too are depictions of these practice constituents.

[30] The ambiguity of 'of' in the phrase 'picture *of* a baseball bat' is discussed in the section below, "Material and Depictionary Scenes."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Just as to hold a depictionary mirror up to a pig is to reveal a pig, so too to hold a depictionary mirror up to a batter striking out is to reveal a batter striking out.

Whether the depictionary practice adds to the identity of the depicted practice constituent or not, there is no room for reasonable debate on the identity of these doubly or singly informed depictions. Is the depiction of a batter on his third futile swing at a good pitch a depiction of a batter striking out, or not? is an absurd question. (Again, just as one can question whether or not the pitch is good, one can also question whether or not the depiction is of a good pitch—it might be shown a hair off the plate. *This* indeterminacy does not affect my point.) And, again, even if one can offer good and compelling reasons why four-strikes-and-then-out baseball would be morally or rationally preferable to our present game, the depiction's identity as a depiction of striking out and as a depiction of the batter being bested by the pitcher is a function of the extant, not the ideal, practice. Again, one *cannot* say, "That is not *really* a depiction of a batter striking out or being bested by the pitcher because it would be ever so much better if four strikes, not three, were out." Again, that would be at best an extremely misleading way of saying that the present practice of baseball should be altered. The predicates, including value predicates, that properly and accurately characterize depictions of doubly or singly informed practice constituents are, like those predicates which characterize the nondepictionary practice constituents, functions of the extant, informing practice or practices, not of a material "practice-free" reality, and not of some other ideal or imaginary practice.

Predicates,[31] like 'is striking out', may apply to both practice constituents and depictions of practice constituents. However, what I call the *transfer* of these predicates from constituent to depiction may occur in one of two ways, which I call *direct* and *prepositional*.

DIRECT AND PREPOSITIONAL TRANSFER OF PREDICATES

A predicate may be said to *transfer directly* from a practice (or, for that matter, from a nonpractice) object, event, or action to a depiction of that object, event, or action, just in case the given predicate applies both to the action, object, or event and its depiction without equivocation or other change in meaning, and the predicate applies to the depiction because it applies to the object, action, or event. For example, the predicate 'is red' applies to a real rose and may also apply, in the same sense of 'is red' to a depiction of the rose, and it

[31] I use the term 'predicate' to refer to the entire expression that attaches to the subject term. 'Predicate', in this sense, includes the copula.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

applies to the depiction because it applies to the rose. In other words, the picture of the rose is red because the rose itself is. Thus, 'is red' is a predicate that transfers directly from the object to the depiction of the object. Or, alternatively, we can say 'is red' is a directly transferring predicate.[32]

On the other hand, there are predicates that transfer from the object, event, or action to a depiction of one of these only with the aid of a preposition such as 'of' or 'about'. A predicate transfers only *prepositionally* if, though it does transfer, a direct transfer of the predicate produces a change in the meaning of the predicate, or the predicate does not apply to the depiction because it applies to that which is depicted. For example, the predicate 'is sweet-smelling' applies to a real rose. But this predicate does not transfer directly to a depiction of the rose. The photograph of the rose is not sweet-smelling. Rather, it is a photograph *of* a sweet-smelling flower. Thus, 'is sweet-smelling' transfers from the rose to the photograph of the rose only prepositionally, or, we can say, 'is sweet-smelling' is a prepositionally transferring predicate. (Of course, any predicate that transfers directly will also transfer prepositionally.)

There are also predicates that might be said not to transfer directly *or* prepositionally from objects, events, or actions, to the depictions of these. (I say "might be said" because one can explain the phenomenon of alleged nontransfer by claiming that the predicate in question fails to transfer because it does not in fact apply to the original object or event either). Our understanding of this mechanism of nontransfer is made easier if we recognize the practice-informed nature of the constituents of the various representational practices (film-making, photography, etc.), a nature neither more nor less practice informed than that of other constituents of other practices. In order to emphasize this practice-informed nature and to explain what seems to be a total lack of predicate transfer in some instances,

[32] What will count as a predicate applying to the depiction *because* it applies to the "original" is itself practice mediated. For example, a photo could be taken of a bitter-tasting rose and, in developing the photo, the photographer might take a bite of the rose and its bitter taste might cause her to spill a bitter-tasting fluid on the photograph. It would then be true to say that the photo of the rose was bitter tasting because the rose itself was bitter tasting; that is, the rose's bitterness was part of the causal and explanatory link that produced the bitter-tasting photo of the rose. Nevertheless, 'is bitter tasting' is not here a directly transferring predicate since the practice of photography has no ideological room for this transfer. (That is, the rules and procedures of photography—which are part of its ideology—do not incorporate the concept of taste, though they might come to do so, at which time 'is bitter tasting' might become a directly transferring predicate.) Thus, part of what determines the existence or absence of direct predicate transfer are the (perhaps changing) ideologies of the involved practices.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I distinguish below between what I call "material" and "depictionary" scenes.

MATERIAL AND DEPICTIONARY SCENES

In one sense of 'of', the question, What is that a depiction of? can mean (using the practice of film-making as the clearest paradigm): What fakery (e.g., ketchup "blood," styrofoam "snow") was used to produce the desired representational effect? What *materially speaking* was before the camera? In this sense of 'of', film footage from a Western may be a picture *of* an actor named John Wayne riding a mechanical horse across a hot, klieg-lit soundstage in southern California. That characterization or identity can be called *the material scene.* In another sense of 'of', the question, What is that a depiction of? means What is the depiction supposed to depict? (What, in my argot, is the practice identity of this constituent?) In this sense of 'of', the answer to that question is: It's a picture of Tex Walker (the character) riding across the snowy plains of Montana on his trusty horse Spike. This is the *depictionary scene.* Clearly, if the material 'of' were what we used in the question, What is that a depiction of? all movies would be movies of actors making movies, which clearly they are not.

This ambiguity creates a problem, or at least a pseudo problem, because predicates that apply to material scenes often do not transfer in any way to depictions of those scenes. For example, if the soundstage in the above-noted scene is very hot because of the klieg lights, the predicate 'is hot' does not transfer, even prepositionally, to the depiction. It is not a depiction of a hot day. It is a depiction of a cold, snowy day.

This ambiguity of 'of' and the choice between providing a material or depictionary description of such scenes is, however, just another (fairly well hidden) example of the same confusion reflected in the Is it really a bat or really a piece of wood? question. The representational arts are *also* practices, and their practice constituents have real identities that are practice informed. The film scene with John Wayne is not "really" a scene of an actor named John Wayne on a mechanical horse any more than a bat is "really" just a wooden stick or a baseball player is "really" just a man in funny pants. The identity of the practice constituent (the scene on film) just *is* its practice identity. There is none more basic or more real. Of course it is a shot of a snowy day. Of course it is a picture of Montana. (Cf. "Of course he is striking out—that's his third swing.") Naturally, *in another practice,* the identity of these constituents changes. But that we already knew. Not that this changed constituent would *look* any different, any more than a bat looks different from a (bat-shaped) piece

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of wood. It is the practice that produces the identity of the constituent. Change the practice, and you change the identity.

A complication arises in identifying the practice constituent in those cases in which the representational practice depicts a constituent of another (additional) practice. For example, in a movie about baseball, the identity of a given depiction is a *twin* function of *both* informing practices—here, baseball and film-making. So what might we have a depiction *of*? An overweight actor pretending to hit a home run with a papier-mâché bat? No, that is the material scene. The practice identity of the depiction (which is to say, its identity) is that it is a depiction of a famous baseball player hitting a home run. What gives the depictionary scene this identity is *both* the practice of film-making and the practice of baseball. That is, the practice of film-making informs the identity of the depiction no less than the practice of baseball. In this practice, a depiction of something made, materially, of papier-mâché is a depiction of a wooden baseball bat, not a depiction of a papier-mâché movie prop.[33] That practice constituents (baseball bats, surgeons' scalpels, KKK crosses) have a practice identity (regardless of their material identity as plaster of paris, tin foil, or cardboard) in fictive settings is no more remarkable than that the English language, spoken in a novel or on film, retains its standard meaning. The sentence, 'He has arrived', spoken on a stage, still means what it does in life, even though no one has arrived

[33] This distinction between material and depictionary scenes is of direct importance in the moral and juridical evaluation of the pornography issue, since it is often said that pornographic portrayals are *really* only of actresses and actors playing scripted parts (tomorrow the porn queen will play Lady Macbeth) and not, say, photos of real rapes or other real crimes against women. But this opinion is the product of a double confusion. It is a confusion because, first, there is a failure to understand the doubly practice-informed nature of these pornographic depictions. A pornographic depiction of a rape is a depiction of a rape, not of an actress pretending to be raped, just as the depiction in the Western film is of Tex riding across Montana on a horse, not of John Wayne on a soundstage in California. A depiction of an actress pretending to be raped is a different scene, though it might not look any different (recall again the bat and the bat-shaped piece of wood). Secondly, those who attempt to ameliorate the possibly severe moral judgment of pornographic depictions also fail to note that, if the stories of the survivors are true, pornographic depictions often have material and depictionary identities that are one and the same. [See for example Linda Marchiano's *Ordeal*, written under her better known name, Linda Lovelace, with Mike McGrady (Seacaucus, N.J.: Citadel Press, 1980)] That is, depictionary scenes of a woman being raped may be produced by filming material scenes of a woman being raped. In some cases, this collapse between material and depictionary scenes is highly contingent (one can easily produce a depictionary scene of a woman being raped without an identical material scene), but, in other cases, the collapse may be all but necessary (How can one produce a photographic depiction of a woman's exposed genitals without an identical material scene?).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and an actor speaks the words without the usual intention. Similarly, a film's practice object, e.g., a cardboard bat, is a baseball bat, not a cardboard movie prop. Like the English language, the bat's identity is its practice identity—and here that identity is a twin function of the practices of baseball and film-making.

DIRECT AND PREPOSITIONAL TRANSFER OF VALUE PREDICATES

A *value* predicate is a predicate that ascribes moral or nonmoral goodness or badness itself (where goodness and badness are inclusive of rightness and wrongness) or ascribes that which is morally or nonmorally good or bad to some practice (or nonpractice) object, event, or action. Some value predicates transfer directly from "original" to depiction, and some transfer only prepositionally. An example of a direct transfer of a value predicate from an object to its depiction would be the predicate 'is beautiful' as applied to a rose. This value predicate transfers directly from the rose to a photograph of the rose, since the predicate may apply to the depiction in the same sense in which it applies to the rose, and may apply to the depiction because it applies to the rose. On the other hand, the predicate 'is a valuable floral specimen', transfers only prepositionally to the depiction of the rose. The depiction is not a valuable floral specimen, rather, it is a depiction *of* a valuable floral specimen.

This direct transfer of value predicates may take place between many practice constituents and their depictions. For example, consider the practice of social etiquette. Within this practice, certain actions shame or disgrace the agent of those actions, whereas others bring the agent social approval or even honor. Mistakenly drinking the water in one's fingerbowl, for example (especially if one slurps it loudly and exclaims, "Great soup!") is an action that would surely disgrace the foolish dinner guest. In describing this incident, we might say, e.g., "Clara was disgraced by this action," or, alternatively and equivalently, "This action disgraces Clara." Does the predicate 'disgraces Clara' transfer, though, to a depiction of this unfortunate event? (Let us say that the host's nephew, Elwood, was videotaping the party.) And, if the predicate, 'disgraces Clara', does transfer, does it transfer directly or prepositionally? Is Elwood's videotape footage of Clara slurping the water in her fingerbowl a depiction that disgraces Clara? Or is it merely a depiction *of* an action that disgraces Clara? (Again, if the predicate transfers directly, it also transfers prepositionally.) It goes without saying that drinking water in one's fingerbowl is a practice action; its meaning and significance are a function of its practice membership. In another practice, it might represent the height of sophistication and good manners, a point irrelevant here. I believe that critical reflection reveals that the predi-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

cate in the above instance transfers directly from the practice action to the depiction of it. Elwood's videotape footage disgraces Clara. The predicate 'disgraces Clara' is used with the same meaning when applied to both action and depiction, and it applies to the depiction because it applies to the action. It is not the case that Clara is literally disgraced by the actual drinking and only metaphorically disgraced by Elwood's videotape.

The same direct predicate transfer would occur if the fingerbowl-slurping episode were part of a movie, and Clara a character in that movie. The character would be shamed or disgraced by her actions as revealed in the movie. Here the predicate transfers from this constituent of the practice of social etiquette as an extant material and conceptual construct, rather than from a particular, real-life fingerbowl slurping. (Transfers, we might say, from slurping as a type, rather than a token, to a token.) There may be no value-predicate transfer from the material scene of actress Jane Smith pretending to drink water in a fake fingerbowl to the depiction on the screen.

Let us consider another example of direct predicate transfer of a value predicate. A baseball player has the fans standing in the bleachers, applauding his record-breaking home run. Given the conventions of the practice of baseball—given this ovation's practice identity—this group applause is an action that honors this player. A photograph, taken by a local newspaper reporter, which depicts the standing ovation, also honors this player. The predicate 'honors this player' transfers directly from the practice constituent (the ovation) to a depiction of it. Again, there is no change in the meaning of the predicate, and the photograph honors the player because the original action honored him.

The following final example of direct predicate transfer of a value predicate is instructive because it illuminates the relationship between the performative power of certain utterances within certain practices (here, "I award you this medal") and the process of predicate transfer.

In an honor-awarding ceremony in the military, the utterances made are performative; that is, the utterances themselves accomplish the task of bestowing honor. The medal wearer is then honored, or continues to be honored, by wearing the medal bestowed in this ceremony. We might say that the "power" of the bestowal of honor moves, in this practice, from the performative utterance to the related object, the medal itself. The soldier so honored now continues to be honored by wearing his medal. A depiction of the soldier wearing his medal also honors the soldier. 'Honors this soldier' is here a directly transferring value predicate. The picture honors the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

soldier in precisely the same sense in which the medal does, and it honors him because the medal does. However, both the medal itself and the photograph honor the soldier in virtue of their relationship to the honor-awarding ceremony of this practice and to the performative utterances of that ceremony.

Generally, as the above examples bear out, the value predicates that directly transfer from practice constituents to depictions are those which have a strong notional component—*not* that we can mentally assign the predicates or not, as we might with equal facility think of a red or of a green triangle—but that the practice's assignment of the predicate to the practice constituent is locked within and constrained by the real and interrelated meanings of practice experience rather than, say, constrained by physical laws or logical impossibilities. Put another way, we can say that these directly transferring value predicates are more closely related to the real social meanings that create our experience than to the material edges of reality, though it *should not be forgotten* that these material edges do have a bearing on this predicate assignment as well (and that the meaning/material distinction is ultimately specious). Reflection and example might reveal these directly transferring value predicates to include 'shames', 'honors', 'degrades', 'defames', 'exalts', 'elevates', and—significantly for the case under examination—'subordinates'.

THE IRRATIONAL ASSIGNMENT OF VALUE PREDICATES TO PRACTICE CONSTITUENTS

Suppose that, as may in fact sometimes be the case, those who receive medals in the military—and that is to say those who are honored—are men who should not, ideally speaking, be honored. (Suppose that they are Nazi soldiers being honored for killing Jewish civilians.) Can we then say, The photo of this Nazi soldier with his medals is not a photo that *really* honors the soldier, because those medals are *really* a sign of dishonor? This claim, like the claim that the baseball player on his third strike is not really striking out, because it would be objectively better if four strikes were out, is a confusion. It is a misleading way of saying that the practice that awards medals for killing Jews should be abolished. We *can* say that the practice, through its constituent (the medal), presents what is evil as good, and that a depiction of the practice constituent also presents what is evil as good. Indeed, it is this axiological transposition that makes a morally unacceptable practice morally unacceptable.[34] In presenting the dishonorable as honorable, the Nazi practice condemns itself. A

[34] The words from Isaiah (5:20–23) are apposite here: "Woe unto you who call evil good, and good evil, who turn darkness into light and light into darkness, who make bitter sweet and sweet bitter. Woe unto you!"

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

photo of a Nazi wearing all his medals for killing Jews is a photo that honors what is evil and dishonorable. The photo honors what it should not because the medals honor what they should not. To say this is not to say that the photo does not *really* honor the soldier, because it does. We could say that those who are honored by practices that honor the dishonorable are thereby and in fact dishonored by being honored in this way, but the locution 'not *really* being honored' remains misleading, and it in fact *makes inexplicable* our proper value judgment of the depiction. If the Nazi is not really being honored by this photo, then why is this picture contemptible? In looking at the photo of the Nazi with all his medals, we do *not* say, "Ah, finally this swine has received his just deserts," but rather, "How terrible that anyone would choose to honor such a man."

In sum, when a practice makes this type of axiological transposition (viz., presenting evil as good or good as evil), to say that the given practice constituent is not *really* good or not *really* evil is a confused way of saying that the practice has performed an axiological transposition, and that no practice should perform such a transposition. We cannot go on to suggest either that *this* abhorrent practice should not have performed the given axiological transposition or to suggest that the transposition be undone. We cannot suggest that *this* practice should not have performed the transposition because such a statement implies that the transposition is inessential to the practice's identity, when that is not in fact the case. Axiological assignments, transposed or not, are part of what sketch out a practice's identity; e.g., part of what gave Nazism its practice identity just was its axiological perversity. And neither can this transposition be undone. The practice's presentation of evil as good or good as evil presents us with a compound meaning for the given constituent which, like all meanings, can never be erased, though, if the transposition is noted within the context of a morally perspicacious society, we may hope that the presentation of this constituent's perverse meaning will fall into disuse.[35]

[35] It seems to be possible for some practice constituents, especially practice objects, to be claimed and re-identified by new practices. For example, the naturally "kinky" hair of Black Americans, along with their very skin color, was seized from its racist presentation as shameful and made into a sign of pride and honor. This change in meaning was made possible, however, not by a mere collective mental act, but by the (partial) abolition of the practice of racism and the birth of the social practice of antiracism. It is significant that other practice constituents of racism, such as lynching and insulting forms of address, could not be transformed or re-identified. There is probably no hard and fast rule determining which practice constituents can be re-identified and which not (Could the Nazi swastika ever become a symbol of Jewish self-affirmation?), and the impossibility of re-identifying most practice constituents is certainly not a logical one.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE MEANING OF 'SUBORDINATION' AS THIS TERM IS USED IN THE PORNOGRAPHY/CIVIL RIGHTS ORDINANCE

Not every action that harms women subordinates women, and not all representations that portray women in a negative light—e.g., representations that make women look foolish or stupid or vain—subordinate women. To subordinate an individual or group of individuals —using the term 'subordinate' as I believe it is used within the context of the Pornography/Civil Rights Ordinance[36]—is to place that person or group of persons socially in the class of those whose intrinsic or inherent moral worth or standing is not of the first rank, and whose rights are thereby of lesser scope, importance, or weight than the rights of others. Such a social placement of women into the class of intrinsic moral inferiors would indeed, as the ordinance also says, represent "a substantial threat to . . . the equality of citizens in the community."[37] The social establishment of a group's intrinsic moral inferiority quite naturally leads to their political oppression, and so close is this relationship between the social stamp of intrinsic moral inferiority and political oppression that it is more illuminating to see these disvalues as twin aspects of a single social coin, rather than as held together by a cause-and-effect relationship.

The social assignment of intrinsic moral inferiority to a group of persons—and, that is to say, their subordination—is quite other than a mere assertion of such inferiority, as would be made by the statement, 'Women are morally inferior to men', or as might be entailed or implied by the assertion, 'Women should be sexually abused by men'.[38] The social assignment of intrinsic moral inferiority to women, the placement of women in the class of intrinsic moral inferiors, the subordination of women, is an actual, empirical, and not merely or essentially linguistic placement, though it is not of course a *physical* placing or moving. One may be assigned a social place without one's body being moved. One may, for example, actually, empirically place or be placed at the bottom of one's graduating class without thereby changing one's spatial-temporal location. One can be subordinated without leaving the house. This fact does not make one's subordination "symbolic" or "metaphorical" any more

[36] See Andrea Dworkin's "A Word People Don't Understand," in *Ms.* (April 1985): 46. Also Catharine MacKinnon in "Coming Apart," *Off Our Backs* (June 1985): 6.

[37] The Pornography/Civil Rights Ordinance, Section One, reprinted in *Ms.* (April 1985): 46.

[38] In "The Minneapolis Ordinance and the FACT Brief," *Women's Review of Books* (May 1986): 8, Rosemary Tong indirectly reveals the irreducibility of such social placement to perlocutionary effect. This particular irreducibility is just a specific instance of the difficulties that follow from the adoption of paradigm shifts discussed below.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

than the above-mentioned scholastic placement is symbolic or metaphorical.

Although the social assignment of intrinsic moral inferiority to women cannot, by its very nature, be father to the *fact* of women's intrinsic or inherent moral inferiority (that would require some effort of patriarchal genetic engineering in which women were, say, turned into tobacco plants), the assignment to the class of intrinsic moral inferiors is nevertheless a real assignment to a real social category. Such a claim for the existence of an actual social category of intrinsic moral inferiority raises many questions beyond the scope of this article. One question would be that of the ontological or metaphysical status of the entity (society) that allegedly has these categories. We might ask whether or not it would be illuminating to attempt a reduction from the paradigm that speaks of a metaphysically real society to one that speaks only of individuals and their beliefs and attitudes. Briefly, I do not believe that this notion of society and the social can be illuminatingly explained by reference only to individuals and their mental states even if, as is surely the case, society is in some sense made up of individuals. Reductionistic schemes or paradigms are unwelcome if we lose predictive and explanatory power through the reduction—which is to say, if the adoption of the reductive scheme causes us to lose the *desired* epistemic access to the phenomenon under examination, the reductionistic scheme cannot be recommended.[39]

Whether or not a paradigm that has no room for a metaphysically real notion of society does in fact cause us to lose desired epistemic access to some phenomenon (here, the phenomenon of subordination) cannot be decided, one way or the other, by pointing to the odd

[39] Of course, whether or not the access is desired is the question. It is axiomatic that change of one's paradigm destroys a certain type of explanatory power, viz., that of the old paradigm. We cannot, for example, explain the moves of a chess player *as chess moves* by describing the movement of the molecules of her body, even though, as is surely the case, the chess player is made up of molecules. The switch from the "Chess Game Paradigm" to the "Physics Paradigm" destroys the possibility of that old explanation being given. In the new paradigm, chess players assume the status of the metaphoric, and to speak of them as real is a reification. Similarly, we cannot explain the phenomenon of the social placement of women into the class of intrinsic moral inferiors *as subordination* by describing the mental states of individuals, even though society is as surely composed of individuals as the chess player is composed of molecules. The switch from the "Social Paradigm" to the "Individual Paradigm" destroys the possibility of the social subordination explanation being given. In the new, reductionistic paradigm, society assumes the status of the metaphoric, and to speak of it as real is a reification.

A good account of the general debate between methodological individualists and holists is given in Simon, pp. 41–55. For a discussion of the relationship between feminism and various forms of individualism, see "The Critique of Individualism" in Jean Grimshaw's *Philosophy and Feminist Thinking* (Minneapolis: Minnesota UP, 1986), pp. 162–187.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

falls from validity which occur if we reason across paradigms. From 'It is certain that the social placement of some group into the class of moral inferiors is unjust' we cannot infer 'It is certain that x-x_n's false belief in y-y_n's intrinsic moral inferiority is unjust'. The inference will not hold even if we drop the modal operator, 'it is certain that', since to believe falsely in another's moral inferiority is not unjust, but only bad. This fall from validity only illustrates the problem of choosing a paradigm, but does not decide the outcome of that choice.[40]

COULD PORNOGRAPHY BE THE SUBORDINATION OF WOMEN?

If to subordinate someone, in the sense explained, is to place that person socially into the class of intrinsic moral inferiors, then 'subordinate' is a value predicate. It is among those predicates which may transfer directly from a practice constituent to a depiction of such a constituent, provided of course that the conditions for such direct transfer are met. Thus, pornographic depictions could subordinate women, if the predicate 'subordinates' transfers without change of meaning from constituent to depiction and if the predicate applies to the depiction because it applies to the practice constituent.[41] As

[40] This paradigm-related fall from validity is, *qua* fall from validity, like that generated by our inability to substitute in extentionally equivalent expressions in certain contexts. (E.g., John believes that Mark Twain wrote *Tom Sawyer*, but it does not follow that he believes that Samuel Clemens wrote *Tom Sawyer*.) But the opacity of beliefs does not place any object of knowledge beyond our reach (John could come to know that Samuel Clemens wrote *Tom Sawyer*), whereas a change of paradigms does do so. Paradigm changes destroy (and create) epistemic objects: "[T]hough the world does not change with a change of paradigm, the scientist afterward works in a different world" [Thomas Kuhn, *The Structure of Scientific Revolutions* (Chicago: University Press, 2d ed., 1970), p. 121.] In the case under discussion, the phenomenon of subordination, and its attendant and obvious injustice, can be lost in a paradigm change.

[41] Since, as I have noted above, the representational arts are themselves practices, it can be the case that the identity of one of their practice constituents (e.g., a picture of a chess game) is doubly practice informed, and thus that identity is not entirely a function of the nonrepresentational practice (here, chess). Although it is clear that some value predicates of *non*practice objects can be generated by the representational practice alone (e.g., a beautiful picture of rotting fruit), it is not clear that nonaesthetic value predicates relating to a nonrepresentational practice constituent (e.g., a home run) can be so generated, much less gainsaid, by the representational practice, especially if one posits the existence of an overlapping ideology that both the representational and nonrepresentational practice share. Specifically, we might ask if it would be possible for pornographic representations to subordinate women even though the identity of the nonrepresentational practice constituents so represented contributed not at all to women's subordination. [If I follow her correctly, Susanne Kappeler's point in *The Pornography of Representation* (Minneapolis: Minnesota UP, 1986) is that the subordination of women may occur through the mechanism of the representational practice alone, regardless of the nonrepresentational practice identity of that which is represented.]

It is certainly the case that the identity of a *depiction* of someone performing practice action x is other than the identity of someone performing practice action x,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

argued above, a predicate may apply to a practice constituent or to a depiction of a practice constituent, even though the assignment of the predicate is a function of an irrationality that has infected the practice. Just as the Nazi's medals honor what are in fact a despicable man's despicable actions—but honor him nonetheless—so constituents in our practice of sexuality and their pornographic depictions may subordinate, in the sense of placing the subordinated into the class of moral inferiors, those who are in fact moral equals—but subordinate them nonetheless. Of course, it is as irrational to subordinate moral equals as it is to honor despicable actions, but we cannot, on the grounds of this irrationality, say either "the medals do not honor" or "the depictions do not subordinate."[42]

So it seems that we have an affirmative answer to the precedential question of our investigation into the conceptual structure of the Dworkin/MacKinnon Ordinance, i.e., an affirmative answer to the question, Could pornography be the subordination of women? The seeming conceptual oddity of the ordinance's definition of pornography as "the graphic sexually explicit subordination of women" is revealed as neither odd nor uniquely generated, given the fairly common phenomenon of direct predicate transfer from practice constituents to their depictions.

A BRIEF OUTLINE OF SOME NORMATIVE ENTAILMENTS OF THE ABOVE ANALYSIS

What is wrong with the reasonable view quoted above is that it fails to allow for direct predicate transfer from practice constituents to their

---

the depiction and the doing being informed by two different practices. It seems to me, however, empirically unlikely, if not impossible, that these identities would not influence each other, though the influence of the nonrepresentational practice is not necessarily the greater. [For an account of general representational to nonrepresentational influence on identity, see Robert Schwartz's "The Power of Pictures," this JOURNAL, LXXXII, 12 (December 1985): 711–720.]

[42] And we could not truthfully make these negating statements about the medals that honor or the depictions that subordinate until such time as these practices have been abandoned as socially real, active practices *and* have either been entirely forgotten, obliterated from human memory, with such total erasure giving their constituents less social meaning than a blade of grass, *or* until they have been entirely replaced by new practices that would generate radically new social meanings. But this entire replacement of one practice by another is problematic. It would require the death of the institutions and ideology of the old practice as well as the complete abandonment of its illegitimate but nevertheless familiar and rewarding (to some) elements. The difficulties here are that, as long as human beings exist, the destruction of practice-generated meaning through social amnesia is a practical impossibility, and the replacement of one practice by another, in that it requires giving up what is perceived by some as familiar and good, is unlikely to occur except through a massive social cataclysm—and the degree of the massiveness of this required cataclysm would be a function of the degree and importance of that perceived good. Given these difficulties, it is more often the case that the "new" practice is not so new after all.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

depictions, a failure which is compounded by the reasonable view's implicit adoption of a reductionistic paradigm which destroys epistemic access to the phenomenon of subordination, as such subordination is referred to within the context of the ordinance. Thus, the reasonable view does not have conceptual room for naming the production or dissemination of pornography as an injustice, but can at most see such actions as producing a state of mere disvalue. Since there is a morally relevant difference between producing a state of mere disvalue and bringing about an injustice, as well as a morally relevant difference between the additional and further harm engendered by producing a state of mere disvalue and the harm engendered by bringing about an injustice, the reasonable view will have different normative entailments from the civil-rights view.

On the civil-rights view, the additional harm related to producing or disseminating pornography (such as the harm represented by sex crimes) is directly and relevantly related to the producer's or disseminator's contribution toward the creation of women as a subordinated class, a contribution and creation which is itself an injustice. On the reasonable view, any further harm that is causally related to producing or disseminating pornography is related to the permissible, if devalued, action of producing or disseminating this merely depictionary material. On the civil-rights view, the additional harm is akin to that harm caused by a bank robber's pistol waving, which causes a frightened customer to die of a heart attack. The robber may be held responsible for this death, even though he did not intend it, because it is a causal consequence of a certain faulty and risk-creating aspect of his action. On the reasonable view, the harm caused by pornography is akin to that harm caused by my rude rebuff of a fellow bank customer's friendly greeting, which rebuff causes him to die of a heart attack. I cannot be held responsible for this death because, even though it is causally related to my faulty action, the faulty aspect of my action (its rudeness) did not create the risk of death or the realization of that risk. Rude actions do not ordinarily create such risks; and, if we find them faulty, that is not *why* we find them faulty. In the case of the bank robber, however, we can say that his pistol-waving action violated the rights of the bank customers to (among other things) personal safety in a public place, and it was just this faulty aspect of his action which created the risk and the realization of that risk.[43] Unjust actions often, though not always, create such additional risks. On the civil-rights view, it is the sexual subordi-

[43] I draw this distinction between that which may and may not be imputed to the agents of faulty actions from Joel Feinberg's "Sua Culpa" in his *Doing and Deserving: Essays in the Theory of Responsibility* (Princeton, N.J.: University Press, 1970), pp. 187–221.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

nation of women that is pornography which grounds and explains the faulty aspect of producing or disseminating it, and this faulty aspect creates an additional and further risk, the risk that members of this subordinated class will become the victims of sex crimes. *Why* the initial act of production or dissemination is faulty (because of its contribution to the creation of a subordinated class of persons) and *how* the additional harm (produced by sex crimes) comes about are linked by the faulty aspect of the original act. Actually to contribute to the creation of a class of subordinated people is concomitantly to create the foreseeable risk that they will be victimized by those who are acting in response and relation to this actual subordination. Contributing to the actual creation of such a subordinated class of people is relevantly different from suggesting or recommending that others engage in certain illicit behavior—although the latter may be within the bounds of just action, the former is not.

On the reasonable view, pornography is merely depictionary material, and, though it may be bad or faulty to produce or disseminate such material, in that the material depicts or recommends the subordination of women, the material does not literally or actually subordinate women, and therefore the producers and disseminators of this material are not bringing about a state of affairs in which women are actually subordinated, but are only bringing about a state of affairs in which women are depicted as subordinated. Therefore, these parties are not responsible, on the reasonable view, for what might happen if they *were* to bring about such a state of affairs of actual subordination, for they have not brought about such a state of affairs. On the reasonable view, mere depictions, as such, do not contribute to the creation of the practice identity of women as a subordinated class; thus, in producing or disseminating such depictions one is not engendering either a state of injustice or even a direct risk for others. Of course, those who hold the reasonable view recognize that agents acting upon the information or recommendations provided by pornographic depictions might harm others. If this occurs, however, the agents of harm are themselves responsible for the harm, not the creators or disseminators of the depictions. One who holds the reasonable view will naturally say that the agents simply should not have acted on the information or recommendations imparted. On the civil-rights view, pornography not only provides information and recommendations, but contributes to the actual and literal creation of women as a subordinated class, a class whose members may then be additionally victimized by those acting in response to this pre-established subordination.

MELINDA VADAS
Oak Ridge, North Carolina
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

Hegel. M. J. INWOOD. London: Routledge & Kegan Paul, 1983. 582 p. $45.00.

The subject of this book is Hegel's system, his account of the "fundamental nature of the universe" (1). The first part, an introduction to the system, discusses Hegel's notion of pure thought as distinguished from ordinary concepts and from perceptions (ch. I), the role of pure thought in his conception of the self (ch. II), and his view of the natural sciences (ch. III). The second part, turning to the problems the system was designed to solve, describes Hegel's general idea of philosophical problems and of the task of philosophy (ch. IV) and then deals with two sets of problems in particular, those of epistemology and those concerning infinity. Hegel's rejection of a Kantian epistemology and the epistemological difficulties of his own system are considered under the first heading (ch. V). Under the second, Hegel's "infinite objects" (freedom, spirit, God—see *Encyclopaedia* § 8) and the possibility of a finite cognition of them are investigated (ch. VI). Hegel's view of the traditional proofs for God's existence and his central metaphysical claim that God is spirit receive particular attention (ch. VII). The third part, "The System," covering the second half of the book, discusses Hegel's conception of logic (ch. VIII), the transition from logic to nature (ch. IX), the metaphysical status of finite things (ch. X), and Hegel's views on freedom and 'Ought' (ch. XI).

The book is an argument with Hegel on these matters. Quoting generously, Inwood points out ambiguities in the texts, formulates Hegel's theses, explains the reasons that Hegel does or could put forward, raises objections, considers defenses. Throughout he wants to know not only what Hegel's system is, but also whether it is true.

He finds that most likely it is not. Chapter I: Hegel failed to give a clear meaning to his talk of pure thoughts. Chapter II: the reasons for his conception of the self are weak. Chapter III: his view of the natural sciences is seriously defective. And so on. Not that Inwood is particularly furious in denouncing his author. With the exception of one "elaborate nonsense" (457), the tone of his criticisms is subdued. But the result is devastating. Virtually none of the ideas considered bears scrutiny.

So one wonders why Inwood wrote the book at all. True, "arguments may be novel and interesting and thus worthy of study, even if they are not valid" (520). The trouble with the arguments he dis-

0022-362X/87/8409/0512$00.50 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

cusses is that they are not even interesting. They are not, because their conclusions are not. Most of the claims in what Inwood takes to be Hegel's system are, from a contemporary point of view, too absurd to merit attention. There is no point in considering, for example, Hegel's grounds for holding that the self is the system of pure thoughts or that the universe is a mind, since most people think that the chance of such statements being true is negligible. Thus there is something dishonest about the book. It purports to be arguing with Hegel in all seriousness, but really nobody can take seriously what the argument is about. Five hundred pages for Hecuba.

"So we may forget about Hegel entirely?"—Not so. The problem lies not with Hegel, but with Inwood's reading of Hegel. Some of the absurd doctrines he claims to find in Hegel are not really there, and plenty of interesting doctrines that are there he fails to find.

For example, Inwood holds (35, 42–45, 88, 308, etc.) that Hegel identified the self with the system of pure thoughts. This reading is based on the following passage (from the preface to the second edition of *Science of Logic*):

> Accordingly, we can then much less suppose that the thought-forms which run through all our conceptions . . . serve us, that we have them in our possession and not rather that they have us in their possession; what remains to us against them, how are *we,* how am *I,* supposed to place myself as what is more universal, *beyond* them, when they themselves are the universal as such? When we place ourselves in a sensation, purpose, interest and feel ourselves confined in it and unfree, then the place into which we can extricate ourselves and withdraw into freedom is this place of self-certainty, of pure abstraction, of thinking (Inwood's translation, p. 35).

Inwood's idea seems to be that by the phrase "what remains to us against them?" Hegel is saying that there is no difference between the forms of thought and ourselves. His argument, turning on the universality we allegedly share with the forms of thought, may be obscure. Still, denying the difference, he must mean that we *are* the thought-forms.

This is a misunderstanding. For one thing, Inwood fails to notice that the second sentence of the quotation contradicts his reading. If I can withdraw into thinking, I *am* not the thinking, let alone the forms of thought. But, leaving that aside, the interpretation misses the point of the first sentence, too. Hegel is arguing in this passage against taking the forms of thought as a mere means to be applied in acquiring knowledge. In this context, "what remains to us against them?" does not mean that we and the forms of thought are the same thing. It means that we are not related to the forms of thought in the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

way the user is related to what he uses; for we cannot understand ourselves without regard to thinking and hence to forms of thought, whereas the idea of someone's using something requires that we have an idea of the user independent of his using this thing. We are not masters of the forms of thought, because they are involved in whatever conception we can reasonably form of ourselves. As Hegel puts it a few lines later, thinking is our very own and innermost activity ("eigenstes, innerlichstes Tun"). This is different from saying that we are the thinking or the forms of thought. If Hegel is asserting anything here, it is a traditional point: man is essentially a thinking being. To be sure, that may be false. But it is not absurd, as the identification of self and pure thoughts is which Inwood finds in the text and duly refutes. However, the real pity is that the absurd thesis distracts Inwood from the important, if puzzling, things Hegel did say about the self. Most strikingly, the idea (from *Phenomenology*) that to be oneself is to be related to others in desire and mutual recognition, an idea as difficult as it proved seminal, goes unnoticed. But more could be made even of the sentences quoted above. Hegel may be read as saying: Granted the traditional point that we are thinking beings—how are we related to our thinking? Do we employ it? But how are we to describe the "employer," independently of his function? Do we really "have" thoughts? What does "have" mean here? What about the thoughts in whose grip *we* are? And so on. Inwood, bent on expounding a system, is insensitive to the skeptic in Hegel. But that may be the role in which Hegel is most helpful for present thinking.

Things are similar with respect to what Inwood considers to be Hegel's central thesis, viz., that the universe as a whole is a mind (1, 246). He ascribes this view to Hegel on the basis of *Encyclopaedia* § 50, where Hegel says that God is spirit ("Geist"). But this does not mean that the universe is a mind. Substitution of 'the universe' for 'God' is problematic, but let it pass. Substitution of 'mind' for 'spirit' is a mistake. Mind is only what in Hegelian terms is called "subjective spirit." To say that the world is a mind suggests that it is a sort of person, a subject; and Inwood endorses this reading (252, 257). So, according to him, Hegel's metaphysical story is that of a mind projecting objects outside it of which to become conscious, gaining self-consciousness in humans, and increasing it through human history, until complete self-consciousness is achieved and history comes to an end. That story cannot be taken seriously these days. Nor is it Hegel's story. He does not invite us to think of the world on the model of a mind. Describing mind as subjective spirit, he suggests rather that what a mind is can be understood only by reference to an intelligibility comprising both mind and world. This, though unclear

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and perhaps false, is yet a candidate for truth worth considering. Again the misreading proves harmful in the consequence. Taking Hegel's "spirit" to be mind only, Inwood pays little attention to what is treated in the philosophy of spirit proper. Hegel on the state and society, on the inferiority of the moral point of view, on the function of art in the modern world, on the philosophical content of Christianity—these are some of the Hegelian contributions that many people still find worth thinking about. A few passing remarks excepted, Inwood does not discuss these matters.

It is probably because he chose the wrong text that Inwood presents Hegel at his most boring. A vast majority of passages referred to come from the *Encyclopedia,* most of these again from the introductory sections 1 to 83. Inwood's preference for this book is understandable. Only here did Hegel actually present the system Inwood is looking for. However, Hegel insists in the prefaces that the *Encyclopedia of Philosophical Sciences in Outline,* as the full title runs, is in fact only an outline of the system for the use of his students. It is not only materially incomplete, so that a lot of the more interesting points are merely touched upon or omitted; it also fails to prove statements it makes, in the way scientific philosophy requires. But, since Inwood, when turning to other texts by Hegel, again prefers passages from prefaces and various introductions, the odd result is that, although he prides himself on arguing with Hegel, he never sets his hand to Hegel's arguments. The body of Hegel's thought, the argument of *Logic* and *Phenomenology,* remains unexplored. The book is not really about Hegel's philosophy; it is about Hegel about his philosophy. This is as disappointing as finding an interview with the coaches on TV when you want to watch the game.

Perhaps Inwood's account lacks interest also because he is out of touch with the literature. We are told, as usual, that "my sparse references to the voluminous secondary sources do not reflect the extent of my indebtedness to them" (5f). In fact, there is reason to think that they do. It is true, the quality of studies on Hegel is by and large fairly low. Still, in the last twenty or thirty years people have brought forward a number of good ideas, of which Inwood fails to make use. To take just one example, Inwood's idea of Hegel as "a metaphysician on the grand scale" (1) is naive now that Michael Theunissen[1] has shown how ambivalent and problematic the relationship is between Hegel's *Logic* and his metaphysics.

RÜDIGER BITTNER

Princeton University

[1] *Sein und Schein: Die kritische Funktion der Hegelschen Logik* (Frankfurt: Suhrkamp, 1978).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Lebenszeit und Weltzeit. HANS BLUMENBERG. Frankfurt am Main: Suhrkamp, 1986. 378 p.

The *Book of Revelation* says of the devil: "he knows that his time is short." With this "one-sentence myth" Blumenberg introduces the long central section of this dense, demanding study of the gap that separates *Lebenszeit* and *Weltzeit,* the necessarily limited time allotted to us and the endless time of a world indifferent to our desires. Are we not like the devil in that we, too, know that time is short? We are given too little time to experience, know, and appropriate what the world has to offer. This lack of time lets us turn to surrogates, myths, and metaphorical substitutes. Impatient, we try to speed things up, to cram more and more into less and less time. But impatience, as René Descartes knew, leads to precipitous, ill-founded judgments. If, as Francis Bacon would have it, truth is the daughter of time, shortness of time is the father of error. Blumenberg goes further when he insists that "shortness of time is the root of evil" (71).

Blumenberg's location of the origin of evil in the experienced disproportion of life-time and world-time recalls Friedrich Nietzsche's discussion of the "spirit of revenge," the "ill will against time and its 'it was'."[1] Unable to accept that every future will some day be past, that we, too, some day will have been, we dream of some paradise not shadowed by the shortness of time, where attempts to realize such dreams are inevitably paid for with a loss of reality.

Yet we find it difficult to renounce such attempts, especially we philosophers. The aura of paradise thus colors Edmund Husserl's construction of the *Lebenswelt,* which, as this so successful coinage suggests, promises to close the gap between life and world and thus to banish oppressive thoughts of times from which we are excluded, raising hopes for a fuller life, a life not estranged from itself by an objectifying rationality which has given us power over nature only to let us lose our place in it. Unfortunately, such constructions cannot quiet concerns about our lack of time. Husserl's life-world gestures toward something from which we are necessarily excluded, no matter how hard we try to fill in what philosophy must leave abstract with images drawn from peasant life, primitive cultures, or the subhistorical everyday. The gap between life-time and world-time is constitutive of human experience.

[1] *Thus Spoke Zarathustra,* Walter Kaufmann, trans., in *The Portable Nietzsche* (New York: Viking, 1954), p. 252.

0022-362X/87/8409/0516$00.50
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar
© 1987 The Journal of Philosophy, Inc.

Digitized by Arya Samaj Foundation Chennai and eGangotri

Blumenberg's study of this gap divides into three parts. Each part, though it casts light on the others, can yet stand by itself. By far the longest of these is the central part, which discusses the nature and history of the ever widening gap between life-time and world-time and tells of some of the strategies meant to close or at least narrow it. Blumenberg here traverses some of the same territory he had covered more fully in the *Genesis of the Copernican World*,[2] although now with a different end in view: ever since the eighteenth century the progress of astronomy has provided a paradigm for those who would understand the nature and progress of scientific understanding or of what we can call "theoretical self-assertion"; Blumenberg appropriates this paradigm, but gives it a special turn by underscoring the time requirements of astronomy and more generally of knowledge. The truth takes more time than is available to us.

Hipparchus already suspected this when, confronted by a new and quickly changing star in Scorpio, wondering whether such phenomena were perhaps more common than had been thought, he decided to count the stars and to leave his catalogue to posterity, which might some day be able to answer the problem he had raised. The same suspicion is communicated by Seneca's question whether the comets, whose unpredictable appearance seemed to deny them a place in the superlunar realm with its regular and therefore calculable motions, might not in fact obey periods of their own, periods so long that a single lifetime would not be sufficient to determine them. In the face of the failure of his predecessors to collect the necessary data, Seneca counseled resignation: human beings should not demand for themselves the possession of the whole of knowledge. The time demands of truth call for an ethos that subordinates the individual's desire to satisfy his demand for knowledge to the ongoing pursuit of knowledge, stretching over generations.

With its concern for individual salvation and its devaluation of time, the Middle Ages left little room for such resignation. Nor is it characteristic of the early modern period. Copernicus thought that enough time had passed to warrant confidence in the observations of the ancients and to allow him to complete astronomy, although, already a hundred years earlier, Regiomontanus had questioned the reliability of such observations and wondered whether it might not become necessary to cast off the burden of the past. Descartes thus was to insist on a new beginning, although he, too, thought his own lifetime sufficient to complete the edifice of science with a definitive morality and a perfected science of medicine which held hope even of a conquest of death. Such hopes were soon to be disappointed.

[2] *Die Genesis der kopernikanischen Welt* (Frankfurt am Main: Suhrkamp, 1975).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The completion of science turned out to take much more time than available to one individual or even one generation. Olaf Römer's discovery of the finite speed of light further helped to separate world-time from life-time. "Light, the dominant metaphor of truth and knowledge, transformed itself into an obstacle to both and led to an isolation of that segment of the world accessible from a terrestrial standpoint" (182).

The institutionalization of science may be understood as a response to the widening gap between life-time and world-time. We find consolation in the thought that work we must leave incompleted will be taken up by those who follow us. Placed between life-time and world-time, historical time shields us from the disturbing thought that all we can accomplish will some day be lost forever. Among the most suggestive passages of this book are those where Blumenberg discusses *memoria,* which offers "something like the *intersubjective retention* of the life-time," as "the center of the confrontation of life-time and world-time" (301).

Blumenberg concludes this central part with a chapter "On the Genetic Phenomenology of World-Time." As this title suggests, Blumenberg leaves no doubt concerning his debt to Husserl. But he also forces us to question all attempts to understand world-time as no more than a deficient mode of life-time. Constitutive of our being in the world is the possibility of a self-transcendence that leads equally to the idea of an absolute subjectivity and to the idea of an absolute time; both radically transcend all lived experience. The pursuit of truth is inseparably bound up with such self-transcendence: "The mundane subject completes itself by making the most difficult of all concessions that can be expected of it; to let *its* world become *the* world, to witness the transformation of *its* life-time as only one of many life-times into world-time and as such alienated from itself" (306). Such insight brings with it the recognition that we are not the measure of all things and that it is just this which gives our existence its meaning. "What phenomenology has to observe and to describe is the process of the maturation of subjectivity as the balance it has to find between resignation and fulfillment, renunciation and demand" (306).

Two enormously suggestive essays which engage Husserl's phenomenology in greater detail frame the main body of the book. The first, "The Life-World Misunderstanding," traces the genesis and evolution of Husserl's "conceptual centaur" (19), which makes its first appearance in his 1924 memorial address in honor of Immanuel Kant's 200th birthday. The misunderstanding named by the title is invited by the way the concept of the life-world emerged, not as representing a clearly circumscribed phenomenon, but in response

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to a variety of concerns that emerged from within Husserl's phenomenology and to challenges from without, especially from Martin Heidegger's analytic of Dasein with its emphasis on finitude and death. The concept is misunderstood by those who understand the turn to the life-world as a turn away from the all too formal constructions of academic philosophers to the real stuff of our lives. Although the life-world is "describable as a sphere of constant presence" (34), that description is itself a construction of something that we have always already left behind. The life-world is a limiting concept, born of the attempt to think the possibility of an existence truly at one with itself, which has to include the thought of the convergence of life-time and world-time.

In the final essay Blumenberg turns to Husserl's conception of the *Urstiftung,* the primal establishment of the theoretical attitude in early Greek thought which figures so prominently in the *Crisis,*[3] where Galileo and with him modern science are criticized for having strayed from this foundation. Against Wilhelm Dilthey's and Heidegger's more radical understanding of the importance of history, the *Urstiftung* was meant to secure the historical continuity of the "European spirit." To do so, what is thus established has to transcend the contingency of the world. Blumenberg speaks of a "a geneticized Platonism" to which "something like a cosmogony of the noetic sphere" would correspond (373). This raises the question of how to understand the genesis of the ideality of the noetic from historical facticity. The old problem of the affinity of the ideal and the factical re-emerges. "This would have to be once more the theme of an unknown '*Urstiftung*': the foundation of an ideality of a higher order, which would make the factical capable of genetic ideality. The phenomenologist is defined by fearlessness before the infinite regress" (374).

Anyone interested in Blumenberg's sympathetic yet profoundly ambiguous relationship to Husserl may well find these two essays the most challenging part of a challenging book. I cannot even begin to do justice here to Blumenberg's often annoyingly terse and yet circuitous observations, but only assert that the reader patient enough to give Blumenberg the necessary time will discover that his questioning treatment of Husserl opens up new directions for phenomenology, where a price to be paid is the renunciation of the hope of ever closing the rift between life-time and world-time, and, that is to say, of ever returning to the life-world.

KARSTEN HARRIES

Yale University

[3] *The Crisis of European Sciences and Transcendental Phenomenology*, David Carr, trans. (Evanston, Ill.: Northwestern UP, 1970).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

The Department of Philosophy of Columbia University is pleased to announce the appointments of David Albert as Assistant Professor, Raymond Geuss as Professor, and Sarah Stebbins as Adjunct Associate Professor, effective July 1, 1987.

The Department of Philosophy of Memphis State University announces its annual Spindel Philosophy Conference, to be held October 9–11, 1987. The topic this year is "Connectionism and the Philosophy of Mind." Speakers include William Bechtel (Georgia State), Robert Cummins (Colorado), George Graham (Alabama/Birmingham), Terence Horgan (Memphis State), David Kirsh (MIT/Artificial Intelligence Lab), Paul Smolensky (Colorado/Computer Science), John Tienson (Memphis State), and Michael Tye (Northern Illinois). Further information may be obtained from the Department of Philosophy, Memphis State University, Memphis, TN 38152.

The Department of Philosophy of Yale University announces a conference on Philosophical Perspectives on Newtonian science, to be held at the British Art Center, Yale University, November 6/7, 1987. The main speakers will be Peter Achinstein (John Hopkins), Michael Friedman (Illinois/Chicago), Errol Harris (Northwestern), J. E. McGuire (Pittsburgh), Lawrence Sklar (Michigan), and Howard Stein (Chicago). Further information may be obtained from Phillip Bricker, Department of Philosophy, Yale University, Box 3650 Yale Station, New Haven, CT 06520.

The Stanford Humanities Center will offer eight external fellowships for 1988/89 intended for scholars and teachers in the humanities, or for those in other fields working on related projects. Fellows will spend the academic year at Stanford, pursuing their own research; they are also expected to devote one-sixth of their time to teaching or otherwise contributing to intellectual life at Stanford. Deadline for applications is December 1, 1987. Application materials and further information may be obtained from Morton Sosna, Associate Director, Stanford Humanities Center, Mariposa House, Stanford University, Stanford, CA 94305-8630.

The Department of Philosophy of Rutgers University is pleased to announce the establishment for at least three years of an annual international competition for the best essay written by an undergraduate majoring in philosophy. The deadline for the competition will be November 1 of each year. Submissions will be judged by members of the faculty of the Department of Philosophy at Rutgers, and should be sent to Regina Mayer, Department of Philosophy, Rutgers University, New Brunswick, N.J. 08903.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIV, NO. 10, OCTOBER 1987

पुस्तकालय
गुरुकुल कांगड़ी विश्वविद्यालय
हरिद्वार

HUSSERL AND FREGE*

MANY composers, painters, and writers show their hand in their earliest works, and the difference between Edmund Husserl and Gottlob Frege shows up in an issue that both addressed early in their careers: the philosophical clarification of number.[1] In his *Philosophy of Arithmetic* (1891),[2] Husserl admits that, in most cases, we refer only symbolically to numbers; using counters, we refer to numbers in their absence, and we manipulate the counters according to rules. But the *sense* of what we are doing when we reckon with numerals, he says, comes from those infrequent cases in which we achieve the presence of numbers themselves, when numbers are genuinely given. This presentation, he claims, occurs only for very small numbers. We collect groups and relate them to one another as more or less. We collect this cluster of marbles and that cluster of pencils, and we call the first group "four" and the second "three." It is this sort of collection and comparison which gives us the sense of what we are doing when we talk about 5,794 and 3,296, even though we could never authentically present either of these two numbers. What happens in a genuinely presented group is that "this, this, and this" are grouped, and taken or presented or intended as one, as a whole, and we are the ones who assemble them into a whole. If we want to give a philosophical clarification of what numbers are, we must involve our activity of collecting.

* To be presented in an APA symposium of the same title, December 28, 1987. Ronald McIntyre will be co-symposiast, and Donn Welton will comment; see this JOURNAL, this issue, 528–535 and 535/6, respectively, for their contributions.

[1] Abbreviations in this paper: CP stands for Frege, *Collected Papers*, Brian McGuinness, ed. (New York: Oxford, 1984); FTL stands for Husserl, *Formal and Transcendental Logic*, Dorion Cairns, trans. (The Hague: Nijhoff, 1969); HI stands for David W. Smith and Ronald McIntyre, *Husserl and Intentionality* (Boston: Reidel, 1982); HF stands for J. N. Mohanty, *Husserl and Frege* (Bloomington, Ind.: University Press, 1982).

[2] *Philosophie der Arithmetik*, Lothan Eley, ed. (The Hague: Nijhoff, 1970).

0022-362X/87/8410/0521$00.80 © 1987 The Journal of Philosophy, Inc.

CC-0. Panini Kanya Maha Vidyalaya Collection. An eGangotri Initiative

Digitized by Arya Samaj Foundation Chennai and eGangotri

Frege saw things differently. For him, the number four does not arise through our collecting; the collecting is done by concepts. In his *Foundations of Arithmetic* (1884),[3] Frege speaks of *die sammelnde Kraft des Begriffs* (§48). For example, the concepts "Horseman of the Apocalypse," "cardinal virtue," and "side of a square" each apply to four individuals. The number four thus belongs to each of these concepts and to all other concepts that are equal to these in their extension. And the *concepts* isolate their instances and collect the groups, *we* do not.

Frege thinks it is psychologistic to say that numbers arise through our activity of collecting. He accused Husserl of mixing the logical and mathematical with the psychological (CP 177–179). Frege wants to work only with the objective relationships. Husserl admits that he turns to our collecting activity, but claims he does so in such a way that the objectivity of mathematics, logic, objects, and states of affairs is not compromised: he studies ourselves precisely in our achievement of objectivity, and he studies objectivity precisely as it shows up to us as objective, as reidentifiable and thus transcendent. It seems to me that what Husserl does is to take on directly the problem once admitted by Frege when he said that, when we grasp a thought, "something comes into view whose nature is no longer mental in the proper sense, namely the thought; and this process is perhaps the most mysterious of all." Frege went on to say, "I should say that this question [of how the process takes place] is still far from being grasped in all its difficulty" (CP 145; see Mohanty, HF 37). What Frege called "this process" is what Husserl calls constitution, the achievement of objectivity, the presenting of the objective to us.

As regards the sides of a square, for example, part of their being four is that they can be taken as a group, as one. They have to *be* four for me or anyone else to take them as such; I do not arbitrarily decree that they will be four and only four; but being four they also present themselves as a group of four, and, for them to so present themselves, I or someone like me must be part of the scene as a dative. It would be incomplete to let the concept do all the gathering; *for whom* does the concept gather?

Frege makes the right distinctions, but leaves the things he distinguishes dangling, decontextualized. He distinguishes correctly between signs, referents, senses, and ideas, and also between thoughts, judgments, assertions, and ideas; but he argues only negatively, by retorsion, to the necessity of these distinctions. He shows *that* thoughts could not be ideas, for then there would be no common science; he shows *that* we could not have only ideas, for then there

[3] J. L. Austin, trans. (Oxford: Blackwell, 1959).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
would be no bearer of ideas and, hence, no ideas either. His arguments are economical and convincing, but incomplete, for surely something positive could be said about *how* the terms he distinguishes become distinguished from one another. And the something that could be said involves an appeal to the sort of activity we carry out when we allow a sign to be sign, when we let a sense be differentiated from a thing, when we permit a concept to be differentiated from an object.

Every so often the presence of our activity intrudes upon Frege's analysis. In "On Sense and Meaning," for example, he raises the question of how we know that a particular name has a referent (CP 157–177). He answers that, when we use a name, it is not our "intent" *(Absicht)* to talk about our ideas; we do, in fact, presuppose reference (CP 161/2). Thus, it is our "intent" that constitutes a sign as referential. Frege only touches on this intent, this *Absicht,* but it is the kind of issue which Husserl would explore lavishly, indeed, which he did explore at the beginning of the first of his *Logical Investigations.*[4] A sign is a sign because someone uses it as a sign with which to refer to something. Frege himself, earlier in the same essay, mentions the difference between a sign functioning as a sign and a sign collapsing back into being a simple thing (CP 157). We have to *achieve* something to make a sign be more than a mere thing. But to recognize that a sign is an achievement, that it is constituted as a sign, does not make the sign dissolve into mere ideas or mental states. It is this cloth (flag) or those marks (words) that *are* the sign, and I can take something as a sign only if it permits itself to be so taken. But, once I do recognize, philosophically, the constituted character of signs, I can say a lot more about what it is to be a sign and what happens when a sign occurs. The sign does not just dangle there anymore without the context of use.

In "Thoughts," Frege also appeals to an intent, an *Absicht,* in the case of pictures (CP 351–372). A picture is more than a mere object, such as a leaf or a stone, and it is made to be more by our intent that it present something (CP 352). (He adds that even an idea serves to present something only through our intent.) In Husserl's terminology, Frege would here be touching on—but not exploiting—the intentional activity that constitutes or achieves a picture as such, and, once again, the fact that we achieve the picture does not dissolve the objectivity of the picture. It does not turn the picture into a mental state.

Frege's occasional appeals to fiction illustrate how his analysis would be improved by a more adequate recognition of our part in

[4] John Findlay, trans. (New York: Humanities, 1970), pp. 269–275.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

what happens. In "Thoughts," he observes that both true and false statements are about things. But sometimes the grammatical subject of our sentence may fail to have a reference, and we may not realize that it has none. In such cases, Frege says we unwittingly and unwillingly slip into fiction (CP 362, 367, 389). Thus, if I mistakenly think there is a man in the car outside the house and say, "The man in the car has a lot of patience," I have wandered into fiction without knowing it. But Frege's analysis is inadequate; no one can enter into fiction without being aware of it. To enter into fiction, a speaker must displace himself from his real context and speak in a fictional context, a fictional world. He must duplicate himself and adopt a fictional voice, in a way analogous to the way we displace ourselves outside our real context when we enter into reverie or imagination. Fiction is not just another way of being straightforwardly in the world, an erroneous way, and we cannot clarify what fiction is by speaking only about words and their sense, their reference, and their truth value. We must also speak about ourselves and our possibilities of displacement.

What shall we say about the interpretation of Husserl's noema and "noematic sense" which has been proposed by Dagfinn Føllesdal,[5] David W. Smith, Ronald McIntyre, and others? I think this interpretation is incorrect, and that it reveals a misunderstanding, not only of the noema, but also of Husserl's idea of intentionality, his transcendental reduction, and his idea of philosophy.[6] I will comment on the interpretation as found in Smith and McIntyre's *Husserl and Intentionality*. I will develop Husserl's ideas as they are found in *Formal and Transcendental Logic*, a work that progresses far beyond *Ideas I* and *Logical Investigations*.

Suppose I am addressing a group of people and at one point I say, "The door at the back of this room is open." The listeners can react in at least four different ways. (1) Some of them may just go along with what I say and directly take the door to be open; they may infer or remark that the room will get cold, or that the door should have been shut by the man who came in late. Their thinking and my thinking are concerned solely with a state of affairs, the door's being open, and perhaps with other states of affairs that are related to it. They and I are "believingly" and straightforwardly involved with things in the world.

[5] "Husserl's Notion of 'Noema'," this JOURNAL, LXVI, 20 (Oct. 16, 1969): 680–687; and HI.

[6] For other criticisms of this interpretation, see especially Richard Holmes, "An Explication of Husserl's Theory of the Noema," *Research in Phenomenology*, V (1975): 143–153; Mohanty, HF; Robert Sokolowski, "Intentional Analysis and the Noema," *Dialectica*, XXXVIII (1984): 113–129; John Drummond, "Frege and Husserl: Another Look at the Issue of Influence," *Husserl Studies*, II (1985): 245–265.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(2) Some listeners may be more critical. They do not not just go along with my statement that the door is open. They do not remain simply and believingly with the state of affairs. Instead, they take the state of affairs, the one I have articulated, in a new way. They take it as proposed by me (FTL §44). They may have their reasons for being critical—they may have seen the door closed a moment ago, and so what I say conflicts with what they take to be the case, or they may know that the door shuts automatically—or they may have no special reason at all; they may simply be rather hesitant to go along with all that they hear. These people execute what I would call a propositional reflection. Their reflection constitutes the proposition as such, however; it is not the case that the proposition was there all along, and that they now only turn to it. They modify the state of affairs by taking it in a new way. By taking the state of affairs as proposed by me, they "propositionalize" the state of affairs. The very domain of the propositional is constituted by a new reflective stance into which we can enter, a new reflective stance which modifies how the world has been taken and articulated. Furthermore, "sense" also belongs to this domain, either as a whole proposition or as part of one; indeed, Husserl calls this domain "the region of sense" (FTL §48). We do not begin with the proposition or the sense as somehow part of our mental acts, nor do we begin with the proposition or the sense as subsisting in some third realm, nor are the proposition and the sense there as mediators between our acts and the objects to which we refer. Just as a mark becomes a sign when we take it in a special way, so a state of affairs becomes a proposition when we take it in a new way, as being proposed by someone. Now these critical listeners, having propositionalized what I said, can then turn around and see whether or not the door is, in fact, open, and, in doing so, they disquote or depropositionalize, and they confirm or disconfirm what I said. It is in this way that propositions are matched with states of affairs.

(3) A third group of people might adopt still another reflective stance, one that Husserl would call psychological. They would look upon both me and my audience as entities in the world, and they would think about our activities, about what we are doing. They would talk about perceptions, communications, responses, attitudes, motivations, and the like; and they would treat such things just as normal worldly phenomena, as occurrences that have their causal networks, their motivational patterns, their psychophysical and neurological implications. It would be difficult for those who adopt this psychological stance to account for the propositions, senses, states of affairs, and verifications or falsifications which are constituted and achieved by those who enter into propositional reflection; those who

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

adopt the psychological stance are prone to take, say, a proposition or a sense as a mere subjective viewpoint, as a psychological episode. They do not have the appropriate categories with which to handle verification, truth, meaning, and the like.

(4) Finally, some members of the audience, the philosophers, would adopt what Husserl calls phenomenological reflection. These people are not just naively caught up in things and speeches, as is the first group we described, nor are they engaged in critical reflection and verification, as is the second group, nor are they concerned with pursuing the empirical science of psychology, as is the third. These people enjoy a peculiar detachment and interest. They recognize everything that occurs in all the other stances we have distinguished, and they merely contemplate and describe all these things for what they are, as ways of being known, intended, and presented. These philosophers also examine the various stances we have distinguished as well as the activities executed in them, and they clarify how one and the same self can enter into all these stances and activities. They would describe, for example, the shift in focus which makes a state of affairs become a proposition and the change in intentionality which constitutes an object as a sign. These philosophers would describe the articulative and presentational activity that constitutes a state of affairs on the basis of a perceived thing. They would also examine what it is to be a speaker, what it is to be a listener, what it is to be a perceiver, a rememberer, and so on. Unlike the psychologists, they would have no trouble providing adequate terms to clarify the cognitional and verificational procedures executed by the people who adopt the first and second stances we have described; indeed, the study of truth and falsity on their own terms is the specialty of these philosophers. The philosophers would even discuss what it is for us to be in the world, and what the world is as a general backdrop and matrix for experience and thought. Because they examine what it is to be in the world, they would, in their special reflective stance, have to take a distance to being in the world; as Husserl states it, they would "suspend" world-belief and "bracket" the world and everything in it.

Where is the noema in all this? The noema belongs to the context set up by the fourth stance we have distinguished, the philosophical stance. When we speak of a noema, we mean any thing, state of affairs, feature, aspect, or dimension, insofar as it is examined in its presentability to us and to what we do; and we mean it insofar as it is examined from the stance adopted by the philosophers. The noema is not a component of an act, and it is not, formally, the proposition or the sense constituted in propositional reflection. Everything and anything can be examined noematically; material bodies, for exam-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
ple, can be studied noematically. The tree–noema is the tree insofar as it is analyzed as a perceptual object, as articulable in categorial activity, as intendible in absence, as rememberable, and so on. Furthermore, there is also a propositional noema, since we can examine, philosophically, the proposition as it presents itself to a speaker, listener, verifier, and so on; we can examine the proposition in its presentability. Even a sense can be examined noematically, since we can philosophically examine it as it can present itself and be intended, and as it can recur in many different expressions and contexts.

Now, Smith and McIntyre, following Føllesdal, take the noema to be an abstract intensional entity which mediates between an act and the object referred to, an entity which explains how the act is intentional and how it refers to its particular referent as presented in a certain way, with certain predicates (HI 153, 156, 160, 195, 202). They compare the noema with Frege's "sense," the "mode of givenness" of a referent. They explicitly differentiate the noema from the object an act intends. Thus, for them, when one carries out the transcendental reduction and enters the transcendental attitude, one turns away from objects and turns toward subjectivity, and one discovers the noema as a mediating entity between act and object. The object itself drops out of focus. But, for Husserl, the objects remain there for us in our new, philosophical stance, only now we see objects in their correlation with ourselves and our intentionality. We can now analyze objects, but we analyze them noematically, as correlated to noeses. We do not turn away from objects to noemata.

The interpretation of Smith and McIntyre misunderstands the noema, because it fails to see that, in philosophical reflection, we focus on the correlation between object and subject, noema and noesis. They take transcendental reflection as focused just on subjectivity in its intentional structure, and they claim that we discover the noema in its mediating role when we adopt this focus; but they do not preserve the object as part of what the focus can cover (HI 106, 189). If they did preserve the object, they would not need the noema as a mediating entity. There would be only the intentional acts and their correlated objects.

Smith and McIntyre want to use the noema as a device that would explain how consciousness becomes intentional. But Husserl's philosophy is not explanatory in this way; it does not provide devices, it merely describes. What it describes are not "hidden causes" such as abstract entities, but obvious accomplishments such as identifying across presence and absence, the constitution of groups and numbers, the intentional activities that constitute signs, pictures, and remembrances, the achievement of categorial forms, and the like.
CC-0.In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

There are other issues between Husserl and Frege which deserve deeper exploration. Husserl's notion of syncategorematicals can be compared with Frege's discussion of concepts as functions and as unsaturated parts of thoughts: Husserl's notion of categorial intuition provides a much richer context, and a fuller philosophical explanation, of what we mean by parts of a thought and the syntax that binds those parts together. Husserl's categorial intuition, and the difference between vague and distinct judging, can be related to Frege's description of what happens when we "grasp" a thought and when we assent to it. Even Frege's controversial claim that the true and the false are referents of sentences raises some interesting Husserlian questions, since it implies that falsehood is somehow also part of the world, that the world is not just the sum of positive objects, not just the totality of what simply is, but that what is not is also somehow there to be referred to when we speak falsely. Also, Husserl's general theme of empty and filled intentions and identity synthesis can provide a context for many of Frege's analyses, including his treatment of identity at the beginning of "On Sense and Meaning." These are some of the provocative issues to be found in the fine grain of the writings of Husserl and Frege.

ROBERT SOKOLOWSKI

Catholic University

HUSSERL AND FREGE*

> To be sure, we still have philosophical congresses. The philosophers meet but, unfortunately, not the philosophies. The philosophies lack the unity of a mental space in which they might exist for and act on one another.[1]

EDMUND Husserl summarized the state of academic philosophy in 1929 with these words; but, from our contemporary perspective, they also sound amazingly prophetic.

Recently, though, there have been efforts to find a "mental space" in which the philosophical insights of Gottlob Frege and Husserl,

* To be presented in an APA symposium of the same title, December 28, 1987. Robert Sokolowski will be co-symposiast, and Donn Welton will comment; see this JOURNAL, this issue, 521–528 and 535/6, respectively, for their contributions.

[1] *Cartesian Mediations*, Dorion Cairns, trans. (The Hague: Nijhoff, 1973), p. 5.

0022-362X/87/8410/0528$00.80 © 1987 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and also perhaps the traditions they fathered, "might exist for and act on one another." That such a space does exist for Frege and the early Husserl is generally recognized. They studied and criticized each other's works, they corresponded, they were both opponents of naturalism, psychologism, and relativism, and they shared interests in the foundations of logic and mathematics. It is a disputed thesis, however, that Husserl's more distinctively phenomenological views also belong to this same mental space. I believe that they do, and that properly situating them there can improve our understanding of both Husserl and Frege. My focus in this paper is on Frege and the early Husserl, emphasizing especially some issues on which they differed. I contend that Husserl's early project, on which he and Frege clashed, also guides the course of his later work. And so I argue that the issues that bind Husserl and Frege to a common mental space are also relevant to a proper understanding of Husserl's phenomenology, especially his notion of noema and its relation to Frege's notion of *Sinn*.[2]

Let us begin with the period up to 1900, when Husserl and Frege were explicitly interested in the same topics, logic and mathematics, and the issue was—or, for Husserl, became—psychologism. As is well known, Husserl's book on the foundations of arithmetic was reviewed by Frege, who roundly criticized it as psychologistic. Whatever the degree of justice in this charge, Husserl's own terminology and his conception of the project of investigating the foundations of arithmetic invited it. Husserl held that a philosophical understanding of arithmetic requires a clarification of the nature of arithmetical *thinking*, of the mental processes that make it possible for us to have thoughts about numbers and their interrelations. He criticized Frege's explanations of concepts for failing to "enable us to reproduce in ourselves those mental processes which are necessary for the construction of the concept."[3] Husserl himself sought to describe the basic concepts involved in primitive mathematical thinking, as manifested in counting small numbers of perceptible objects, and then to show how the concept of number itself is derived from reflection on such primitive acts of counting. He characterized all

[2] For other relevant discussions, see: Dagfinn Føllesdal, *Husserl und Frege* (Oslo: Aschehoug, 1958); Michael Sukale, *Comparative Studies in Phenomenology* (The Hague: Nijhoff, 1976); Tyler Burge, "Sinning Against Frege," *Philosophical Review*, LXXXVIII (1979): 398–432; Hans Sluga, *Gottlob Frege* (London: Routledge & Kegan Paul, 1980); J. N. Mohanty, *Husserl and Frege* (Bloomington, Ind.: University Press, 1982); David W. Smith and Ronald McIntyre, *Husserl and Intentionality* (Boston: Reidel, 1982); and Dallas Willard, *Logic and the Objectivity of Knowledge* (Athens, Ohio: University Press, 1984).

[3] Edo Pivcevic, "Husserl versus Frege," *Mind*, LXXVI (1967): 162.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

these concepts, including the concept of number, as "contents" of presentations (or acts), and he characterized the enterprise as an investigation of the "psychological origin" of the concept of number. Frege's response, of course, was that Husserl's project conflated concepts, which are objective, and presentations, which are subjective, and so rendered numbers themselves subjective and private affairs of individual thinkers.

Frege's review highlighted a problem for Husserl's conviction that the ultimate concern of philosophy is the analysis of experience: that very conviction threatens to turn philosophy into a study of the subjective, and so to conflict with an objective and nonrelativistic account of knowledge and of things known. However the review may have influenced Husserl's actual views, his subsequent writings are much more sensitive to the issues Frege raised. But there is an important sense in which Husserl never abandons the kind of project which Frege criticized. While agreeing with Frege that concepts and meanings are objective and intersubjective, Husserl never relinquishes his view that they are *contents* of acts. For Husserl, the analysis of concepts remains tantamount to the analysis of experiences. What Husserl does come to see, with or without Frege's help, is the need to develop a nonpsychological notion of content more adequate to his philosophical purposes. Husserl's later work can be seen as increasingly refined attempts to clarify his initial project, to distinguish it from psychology as a new discipline called "phenomenology," and to extend it to the investigation of experiences of objects of all kinds.[4]

The first major revision of this project is Husserl's *Logical Investigations* (1900),[5] which includes an explicit discussion of linguistic sense. Husserl articulates there a number of theses with which Frege could agree. Senses are distinct from referents and from subjective mental processes. Senses are "ideal," nontemporal and nonspatial, entities. Having a sense is essential to any meaningful expression; having a referent is not. Expressions with different senses may have the same referent. And the sense of an expression determines its referent. But there are also differences in their views; the most significant ones, for our purposes, concern the "ideality" of senses and the relation of senses to acts. For Husserl, senses are *universals* that acts *instantiate;* for Frege, they are *ideal particulars* that acts *apprehend* or "grasp."

[4] See Husserl's reiteration of this project in his *Phenomenological Psychology*, John Scanlon, trans. (The Hague: Nijhoff, 1977), pp. 14–33.
[5] John Findlay, trans. (New York: Humanities, 1970). Hereafter abbreviated LI.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Let us consider Frege's views first: there are two conflicting tendencies in what Frege says about the relation of senses to acts, and Husserl's views represent one way of resolving this conflict. On the one hand, Frege takes great pains to deny any ontological relation between senses and acts. Acts exist in the realm of the mental, which he equates with the subjective and private, while senses—though "ideal" rather than spatial or temporal—are just as objective and intersubjective as physical objects. Furthermore, Frege sometimes suggests, when senses do enter into relations with acts, they do so in the same sort of way that physical objects do, i.e., by being *objects* of acts of appropriate sorts. Physical objects can be perceived; "thoughts" can be apprehended and judgments can be made about them. When a thought is apprehended, Frege says, "something in [the thinker's] consciousness must be *aimed at* the thought."[6] And a thought or sense remains just as "external" to an act of apprehending it as a physical object does to an act of perceiving it:

> The expression 'apprehend' is as metaphorical as 'content of consciousness'. . . . What I hold in my hand can certainly be regarded as the content of my hand but is all the same the content of my hand in a quite different way from the bones and muscles of which it is made and their tensions, and is much more extraneous to it than they are (*ibid*, p. 530, n. 7).

And, of course, thoughts and other senses can be referred to just as any other objects can be. On the other hand, Frege sometimes suggests a more intimate relation of senses to acts. Although senses *can* be referred to, that is not their customary semantic role: expressions more frequently refer to ordinary sorts of objects, and the fundamental role of senses is to direct a reader's or hearer's thinking, not to the thought or sense itself, but to these customary referents. In this spirit, Frege relates the sense of an expression to a "mode of presentation" of its referent. And his attempts to show differences in sense among expressions always turn on differences in the cognitive information about the referent that those expressions would ordinarily convey. These aspects of Frege's discussion suggest that senses do not simply play the role of objects in acts in which *senses themselves* are apprehended, but that they also play some role in acts in which other objects, the *customary referents* of expressions, are experienced.

Frege leaves it completely unclear how senses play this second role. They certainly do not do so simply by being apprehended as objects,

[6] "The Thought," A. M. and Marcelle Quinton, trans., in E. D. Klemke, ed., *Essays on Frege* (Urbana, Ill.: University Press, 1968), p. 531.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

for to apprehend a thought is not in itself to know or pass judgment on a truth value, and one can grasp the sense of a singular term without having a presentation of its referent. Nor does it seem plausible to suppose that we intend ordinary objects only by intermediately apprehending a sense and then passing to the object by means of some second act, say an inference or judgment. Thus, it seems that, in addition to being apprehensible objects, senses must also be related in some more intimate fashion to acts in which objects other than senses are intended. For to play the second, and I think primary, role Frege gives to senses, a sense must be "in" an act in a way more like bones and muscles are in the hand: not in it as what it grasps, but as that by virtue of which it grasps what it does. Husserl's explicit concern is to relate senses to acts in this second way, while at the same time preserving their objectivity.

Like it or not, Husserl's account in LI is much clearer on this issue than Frege's. Although linguistic senses can be apprehended in acts that take them as objects, that is not their primary relation to acts. Primarily, senses are *contents* of acts in which ordinary sorts of objects are intended, and linguistic expressions achieve their semantic properties only by virtue of their relation to such intentional presentations. Language is a medium for publicly "expressing our thoughts" and so enabling others to consider the very same object that we are considering. And this is possible because the *sense* we linguistically express is the "intentional *content*" of the underlying act; it is not the *object* of that act, not something that we apprehend or otherwise intend in the act, but that "in" the act which makes it a presentation of a certain object in a certain way. Husserl calls this element of content the act's "*Auffassungssinn.*" As content of the act, it gives the act its intentional character; and, as expressed sense, it gives the expression its referential character, its character of referring to a certain object in a certain way. Thus, in Husserl, the relation of sense to "mode of presentation" is explicit.

But, if the sense of a referring expression is tied to a mode of presentation of the referent by being the *content* of an act that presents the referent in that "mode," has Husserl not lost the other tendency in Frege's discussion of sense? Has Husserl not construed senses in some subjective "psychological" way, and so simply given up on their objectivity and given in to psychologism? The answer lies in a distinction Husserl had come to make even prior to LI: a distinction between the "real" and the "ideal" content of an act. Real content is specific and unique to a particular act, while ideal content is independent of that act and shareable by different acts, even acts

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
of different persons. The sense of an expression is the expressed *ideal* content of an act, and so is not something subjective or private. If this sounds mysterious, Husserl's solution to the mystery is marvelously simple. *Universals* are ideal entities, shareable by the different individuals that instantiate them, and, on Husserl's view, they are ontologically independent of those individuals. The ideal content of an act is a universal, Husserl says, the act's property of presenting a certain object in a certain way; whereas the real content of the act consists of those temporal constituents which literally make up the act and in which that property is instantiated.

LI provides an elegant resolution of the tension in Frege's characterization of senses and in Husserl's own earlier thought. As universals, senses are just as objective, just as ontologically independent of particular acts and particular "thinkers," as Frege's senses. Also, like Frege's senses, they can themselves be apprehended in appropriate acts. (For Husserl, though apparently not for Frege, this apprehension is a type of "intuition.") But, as universals rather than ideal particulars, Husserl's senses also do closer commerce with acts than that: they are *instantiated* in ordinary acts where ordinary objects, not these senses themselves, are intended.

Husserl himself was nonetheless not satisfied with his account of senses, or ideal contents, as act universals. That account yields, in effect, an "adverbial" theory of intentionality (and reference). An act's intending a particular object in a certain way is due, not to there being some relation between the act and an object, but to the act's having a certain nonrelational property: the property of intending such–and–such object in such–and–such way. By the time of *Ideas* (1913),[7] Husserl has replaced this adverbial account with a more detailed explication of how an act comes to have this property: an act is intentional by virtue of a relation between the act and the act's "noema." But this does not mean which Husserl has abandoned all he said about intentionality in LI. Like the intentional content of an act in LI, an act's noema is an "ideal" entity which Husserl characterizes as a sense or Sinn. [Strictly speaking, Husserl calls just one component of the noema—the component that determines which object an act intends and which properties it is intended as having—a "Sinn" or "noematic Sinn" (see esp. §§88, 89, and 91), but he characterizes the whole noema as a sense of another kind, a "Satz"

[7] *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie,* Erstes Buch (The Hague: Nijhoff, 1950). Translations from this work are mine; cf. *Ideas,* First Book, Fred Kersten, trans. (The Hague: Nijhoff, 1982).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(§133).] Thus, as in LI, an act is intentional by virtue of its relation to a sense or Sinn. What has changed is Husserl's conception of the kind of ideal entities senses are: noemata and noematic Sinne are not universals, not properties of acts, but ideal particulars. In this respect, Husserl's conception of sense has moved closer to Frege's. And noematic Sinne are more "Fregean" in another way as well: they are complex structures of senses organized in various syntactic (Husserl says "categorial") patterns (see esp. §§130/1). For Husserl, it is by investigating the complexities of these sense structures that we uncover the complexities in the ways objects are presented to consciousness.

The "Fregean" turn in Husserl's conception of sense necessitates a change in the way senses are related to acts. Since noematic Sinne are not properties of acts, the relation of a Sinn to the act that "has" it cannot be the relation of instantiation. How, then, *are* noemata, and specifically noematic Sinne, related to acts? Are noemata only related to acts as objects toward which acts may be directed? Some interpreters say so. But those interpreters invariably play down Husserl's characterization of noemata as senses, the continuity he sees between this notion of act–sense and the notion of linguistic sense, and the idea that Husserl's notion of sense is interestingly similar to Frege's. And no wonder. For noemata must then either be the objects that our ordinary acts intend or be intermediate objects, the intending of which somehow directs our acts to ordinary objects. To suppose that senses of a Fregean sort play either of these roles in intentionality seems quite bad phenomenology.

But, now, although I believe there are close similarities between Frege's and Husserl's notions of Sinn, what Frege called "aiming at" a Sinn is not the only, or the primary, way that Sinne enter into relations with acts on Husserl's conception of Sinn. Noematic Sinne are not the objects, but the *ideal contents* of ordinary acts in which ordinary objects are intended. On Husserl's conception of content in *Ideas,* the "real content" of an act includes what he now calls a "noesis": it comprises the "phases of experience" which "bear in themselves the specific character of intentionality" (§85, p. 208). The noesis, Husserl says, is a "Sinn–bestowing stratum" in the act. The Sinn itself, the noematic Sinn, is related to this real noetic content, neither as a property that it instantiates nor as an object that it is directed toward, but as the Sinn it "bestows." Husserl introduces the noema and distinguishes it from the noesis in §88 under the heading "Real and Intentional Components of Experience," and he refers the reader to discussions of sense and content in LI. §97 of *Ideas* is entitled "The Hyletic and Noetic Phases as Real, the Noe-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
matic as Non-Real, Phases of Experience." In §129, Husserl again correlates the noesis–noema distinction with the discussion of content in LI, and, in §133, he explicitly relates the noema to the earlier concept of ideal content. Husserl distinguishes both the noesis and the noema, as "components" or "phases" of experience, from the object intended in an act, and he says that both are found by reflecting on the act itself (see esp. §§87–90).

Husserl's new conception of intentional content as an ideal particular does, of course, present a problem: the problem of just how real and ideal content are related. To call it the relation of "bestowal" is not in itself very illuminating, but Husserl may simply have thought this relation *sui generis* and so not explicable in more familiar terms. One thing is clear, though: if noemata are ideal contents, then a noema is not an object apprehended or otherwise intended in the act whose content it is. In Frege's metaphor, it is not like the object grasped by the hand, but like the structure of the hand which is necessary for its grasping whatever it does. Husserl's project prior to *Ideas* was to find a way to tie sense to content, and meaning analysis to act analysis, in just this intimate fashion. *Ideas* continues that project.

RONALD MCINTYRE

California State University/Northridge

FREGE AND HUSSERL ON SENSE*

There seem to be three crucial differences between the analyses of Robert Sokolowski and Ronald McIntyre and each of them has to do with their interpretations of Husserl, not of Frege:

(1) While McIntyre, following *Ideas I,* treats senses—viewed as ideal, intensional entities—as features of any and all intentional acts (first level) discovered by a reflection on them (second level), Sokolowski, on the basis of *Formal and Transcendental Logic,* rejects this interpretation, and argues that senses are brought about only by a new reflective stance (second level) to straightforward statements about facts offered or taken in a doxic mode (first level). Thus, senses cannot be parts of mental acts, for they come about only by taking in a special way the objects or referents of prior intentional acts.

* Abstract of a paper to be presented in an APA symposium on Husserl and Frege, December 28, 1987, commenting on papers by Robert Sokolowski and Ronald McIntyre, this JOURNAL, this issue, 521–528 and 528–535, respectively.

0022-362X/87/8410/0535$00.50 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(2) McIntyre argues that noemata, a term Husserl introduced in *Ideas I,* are to be identified with senses, understood as components of intentional acts, and to be distinguished from objects, understood as the referents of such acts. Sokolowski rejects this identification by further arguing that the reflection through which we come to identify senses (second level) and psychological activities (third level) is different from philosophical phenomenological reflection proper (fourth level). In the latter, we examine all the other levels in terms of the different kinds of objects known or presented, and in terms of the different acts or activities correlative to them. The noema is found only in this context; Sokolowski identifies it neither as a component of an act nor as a mediating entity between act and object, but as any object, including facts and even senses, "insofar as it is examined in its presentability to us" (p. 526).

(3) While intentionality is dyadic for Sokolowski, consisting only of intentional acts and their correlated objects, intentionality is triadic in structure for McIntyre, comprised of acts, senses (as the manner in which objects are taken or given), and objects.

I will look for a way out of this deadlock by (a) returning to a little known passage in which Frege actually attempts to clarify what is meant by the *Sinn* as opposed to the *Bedeutung* of a concept-word; (b) showing that Frege intended his theory of sense and propositions to apply to scientific discourse, not ordinary talk; (c) suggesting that Husserl's interpretation of noema around the time of *Formal and Transcendental Logic* comes to see it as a family of notions, and thus it cannot be reduced to either (i) the object correlate of intentional acts reflected upon philosophically, or (ii) an abstract intensional entity.

DONN WELTON

State University of New York/Stony Brook

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COGNITIVE SCIENCE AND METAPHYSICS*

COGNITIVE science has had substantial impact on various segments of philosophy. Its greatest influence has been on the philosophy of mind and allied portions of the philosophy of language. I shall not try to trace these significant and ever–widening set of connections. A more recent development is the interface between cognitive science and epistemology. Now there are writers like W. V. Quine who simply identify epistemology with the psychology of belief formation. They seek to replace the older questions of epistemology—questions about skepticism or about the nature and structure of knowledge and justification—with questions about the mechanisms responsible for beliefs (or verbal utterances). On the Quinean view, the connection between epistemology and cognitive science is straightforward. Cognitive science (or its predecessor, behavioral psychology) simply *is* epistemology. I do not endorse this replacement thesis.[1] I view epistemology as primarily a normative discipline. One task of epistemology is to formulate appropriate standards of epistemic appraisal, and a second is to see whether and how these standards can be met. But some of the relevant standards are standards for assessing basic cognitive processes. It therefore turns out that cognitive science indeed has an important connection with epistemology. Only cognitive science can specify what our basic cognitive processes are, and only with its help can we determine which of these processes meet the relevant (mostly truth–linked) standards, or meet them to what degree.[2]

In this paper, however, I shall not pursue the relation between cognitive science and epistemology. I want to explore the possible connections between cognitive science and a different branch of philosophy: metaphysics. Of course, on one philosphical taxonomy, metaphysics includes the philosophy of mind. So all contributions that cognitive science might make to philosophy of mind would equally be contributions to metaphysics. But I shall bracket that portion of metaphysics.

* To be presented in an APA symposium on Epistemology and Philosophy of Mind, December 28, 1987. Patricia Smith Churchland will be co-symposiast, and George Bealer will comment; see this JOURNAL, this issue, 544–553 and 553–555, respectively, for their contributions.

I am greatly indebted to Stephen Schiffer and Holly Smith, as well as Vann McGee, Ray Jackendoff, Julia Annas, and David Owen, for valuable comments on earlier drafts or ancestors of this paper.

[1] The term 'replacement thesis' is used by Hilary Kornblith, in his introduction to *Naturalizing Epistemology* (Cambridge, Mass.: MIT Press, 1985).

[2] See my *Epistemology and Cognition* (Cambridge, Mass.: Harvard, 1986).

0022-362X/87/8410/0537$00.80 CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar ©1987 The Journal of Philosophy, Inc.

Digitized by Arya Samaj Foundation Chennai and eGangotri

There is another way of viewing metaphysics as offering ready entree to cognitive science which also will not be my primary focus (although my material will certainly engage it). This is a conception that P. F. Strawson[3] calls "descriptive metaphysics," a field he characterizes as trying to describe "the actual structure of our thought about the world." If this is metaphysics, cognitive science certainly seems to have *much* to contribute. Many practitioners in different branches of cognitive science—linguistics, cognitive and developmental psychology, cognitive anthropology[4]—view themselves as trying to identify the basic "ontological categories" in terms of which the mind conceptualizes the world.

There is another conception of metaphysics, however, which Strawson calls "revisionary metaphysics" and which has a different focus. It tries to identify the *real* constituents of the world, the ones to which we *ought* to be ontologically committed, whether or not they coincide with our naive or commonsense conceptual scheme. What I want to begin to explore here is whether cognitive science can make contributions to metaphysics so conceived.

I.

A natural approach supporting such a contribution is already present in Hume and Kant, or at least in popular interpretations of them. A familiar line of argumentation goes roughly as follows. Philosophical reflection reveals that certain parts of our conceptual scheme (e.g., space, time, unity, cause, moral value) are really the handiwork of the mind. Therefore, contrary to naive common sense, these are not parts of the "objective" (in Kant's case, transcendental) world. Our initial picture of the world needs revision. A minor amendment to this form of argument could yield a scenario in which cognitive science plays a role. Although Kant thought that the mind's contributions must be identified a priori, contemporary thinking commonly maintains that mental constraints can only be ascertained by familiar scientific methods. This is the mission of cognitive science. Once such constraints are located, metaphysical inferences may be drawn as before. If and when elements of our "manifest image" (to use Wilfrid Sellars' terminology) are found to be products of mental machinery, we may feel forced to transform our manifest image into a more respectable, "scientific" image.

[3] *Individuals* (London: Methuen, 1959), p. 9.

[4] For example, Ray Jackendoff, *Semantics and Cognition* (Cambridge, Mass.: MIT Press, 1983); Eleanor Rosch and Barbara Lloyd, eds., *Cognition and Categorization* (Hillsdale, N.J.: Lawrence Erlbaum, 1978); George Lakoff, *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind* (Chicago: University Press, 1987); Frank Keil, *Semantic and Conceptual Development: An Ontological Perspective* (Cambridge, Mass.: Harvard, 1979); and Susan Carey, *Conceptual Change in Childhood* (Cambridge, Mass.: MIT Press, 1985).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Various materials from research on the sensory systems might potentially be invoked for the purposes of revisionary metaphysics. Facts about color categorization, contour enhancement, or categorical perception might be adduced in support of the claim that the physical world is not (quite) as we perceive it. I shall select, however, a different area of cognitive inquiry to illustrate and then evaluate the kind of maneuver I have in mind. This inquiry concerns the principles used by the mind to structure the world into units or unities. The choice of the example is not dictated by the most recent or revolutionary developments in cognitive science, but rather by the centrality of the topic in our metaphysical scheme.

II.

Apparently universal principles of grouping or unitizing were originally proposed by the Gestalt psychologists. Most contemporary cognitive scientists still see these principles as fundamentally correct. I shall draw here on the formulation and elaboration of (some of) the principles by Ray Jackendoff.[5] Although the domain of these principles was originally confined to perception, there is evidence that warrants extending the sphere of application.

One of the Gestalt perceptual principles is that of *proximity,* illustrated in Figure 1. (1a) is most naturally seen as three circles to the left of two, (1b) as two circles to the left of three.

a. b.

○○○ ○○ ○○ ○○○

Figure 1.

The principle behind this grouping evidently involves relative distance: circles closer together tend to form a visual group. A second perceptual principle is one of *similarity:* elements more similar in internal structure tend to be grouped together. This is illustrated in (2a) and (2b), where circles resembling one another in size tend to be grouped together, in the absence of proximity differences.

a. b.

○○○○○ ○○○○○

Figure 2.

The principles of proximity and similarity are both graded conditions. The effect of proximity can be enhanced by increasing dispar-

[5] *Semantics and Cognition,* especially chs. 3 and 8. For a selected survey of other psychological work on grouping and cognitive constraints, see my *Epistemology and Cognition,* ch. 11.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ities of distance and weakened by reducing them. Analogously, greater relative similarity enhances the appeal of grouping elements together. When proximity and similarity both apply, they can either reinforce each other, resulting in stronger grouping intuitions, or conflict, yielding weak or ambiguous judgments. Different people have strikingly uniform reactions to the same examples, however, whether the cases elicit firm or ambiguous intuitions.

Gestalt principles apply to the temporal as well as the spatial dimension, a point readily illustrated with rhythms. Moreover, the effect of temporal grouping is independent of sensory modality, as can be observed by presenting a given rhythm as beats on a drum, as flashing lights, or as touches of a finger on the back of your neck. Fred Lerdahl and Jackendoff[6] have shown that variants of the Gestalt grouping rules apply in music, where listeners and performers group certain successions of notes (but not others) into a single phrase, certain phrases into a single passage, and so on.

It is not necessary that objects be (wholly) in the perceptual *field* in order for a Gestalt rule to apply. Rather, the mind "reads" the pertinent features into the world. This is especially vivid in the principle of *good continuation,* in both spatial and temporal spheres. Figure 3 represents a range of cases where the mind may or may not see continuity among the shaded regions. In (3a) the shaded regions are quite naturally seen as a single, partly occluded thing, illustrating very good continuity. (3g) represents the opposite extreme. The intermediate cases yield a gradation of judgments, progressing from fairly strong unification in (b) and (c) to quite weak in (e) and (f).

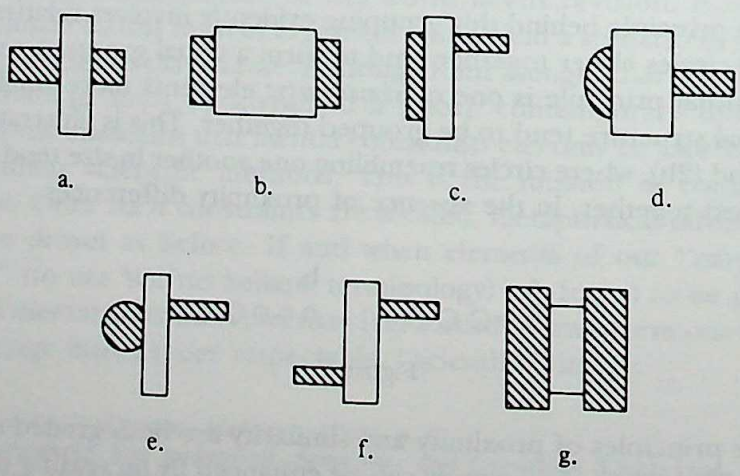


Figure 3.

[6] *A Generative Theory of Tonal Music* (Cambridge, Mass.: MIT Press, 1983).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

As this figure indicates, Gestalt principles are *creative,* in the sense that one makes applications of continuity even in previously unencountered cases. Furthermore, the operative domain of these principles does not seem to be confined to perception. For example, Stephen Kosslyn[7] postulates their operation in imagery, and several psychologists provide evidence that memory encodings of shapes employ partitions derived from Gestalt rules. I shall make another, very natural extension, viz., that the principles govern our unity intuitions in the sphere of "conceptual imagination" as well as perception and memory.

I suggest that Gestalt principles underlie and shape our spatial and cross–temporal "entification" practices, our propensity to view certain sets of spatial elements as parts of one and the same physical object and certain sets of time slices as stages of one and the same continuant. I shall call them principles of unification or unity principles.[8] At least two other philosophers have also drawn attention to the Gestalt basis of unification principles: Quine and Eli Hirsch.[9] Like them, I shall assume that there are strong innate, psychological constraints on "natural" entification. There are complications, however, which I shall not address. The principles I have been discussing are sortal-independent. But it seems clear that we have different principles for unifying or individuating tables and pieces of wood, or persons and brains. How the general psychological principles are related to sortal–specific principles I shall not attempt to say. I do assume, though, that the choice of (linguistic) sortals is itself sharply constrained by human psychology. (Perhaps I should stress that the phenomena in question are just being used as *illustrative* of a metaphysical stratagem. So it is innocuous for me to make certain assumptions, though they may not in fact hold in detail.)

The "partiality" of our unifying principles is brought out clearly by both Quine and Hirsch. Gerrymandered sets of dispersed particles or stages do not count as natural objects. Hirsch describes a unity scheme, called Contacti, in which persons "switch identities" during any interval where they are in contact. Similarly, we can describe a unity scheme that assembles Tuesday stages of Quine, Wednesday stages of Hirsch, and so on. Obviously, no product of these schemes is naturally regarded as a genuine object.

[7] *Image and Mind* (Cambridge, Mass.: Harvard, 1980), pp. 187, 263. Some discussion of work on memory encodings is in my *Epistemology and Cognition,* pp. 230–233.

[8] See John Perry, "Can the Self Divide?" this JOURNAL, LXIX, 16 (September 7, 1972): 463–488.

[9] *The Roots of Reference* (La Salle, Ill.: Open Court, 1973), pp. 54 ff.; and *The Concept of Identity* (New York: Oxford, 1982), especially ch. 8.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

III.

Let me now sketch how this material might be utilized for revisionary metaphysical purposes. The idea is essentially that of Hume's treatment of causation and morals. In connection with causation, he speaks of the mind's propensity to "spread itself on external objects";[10] and, in connection with morals, he speaks of the mind "gilding and staining" all natural objects with colors borrowed from internal sentiment.[11] To use Simon Blackburn's phrase, Hume sees causes and moral values as "projected" onto the world, though they are not "objectively" there.[12] Similarly, some sort of projectivist conclusion might be inferred from the psychological research on unity principles.

How might we best describe the projectivist's view? Consider the kinds of things which satisfy one of the gerrymandered unity schemes, e.g., the scheme that assembles a Tuesday stage of Quine, followed by a Wednesday stage of Hirsch, and so on. Let us call this a "shmerson" scheme. Now our ordinary thought, the projectivist says, is disposed to take a dismissive ontological attitude toward the shmerson scheme. On the one hand, it does not spontaneously think of the shmerson scheme or similar schemes at all. Furthermore, if presented with the shmerson scheme, it is inclined to reject it. True enough, a philosopher might get people to admit that, since the predicate 'shmerson' can be given satisfaction conditions and since there are sets of person stages which meet these conditions, there apparently *are* shmersons. On the other hand, the ordinary person is strongly impelled to add, "But, nonethless, shmersons are not real *things;* they are not real *objects.*" (This is similar to people's initial, and persisting, attitude to Goodman's 'grue' and 'bleen'. These are just gimmicky, nonkosher predicates, people may well say; they do not express genuine properties.) By contrast, people have no such dismissive impulse toward persons (or 'blue' and 'green'). Now the projectivist expresses these facts in the following way. Common sense supposes that there is a property of *entitivity,* such that persons, dogs, etc., have this property, but shmersons, shmogs, etc., lack this property. (Or perhaps what should be attributed to common sense is the view that certain *unity schemes* have entitivity while others lack it.) The projectivist next proceeds to charge common sense with *error,* much as J. L. Mackie[13] charges ordinary moral

[10] *A Treatise of Human Nature,* Book I, Part 1, Section 14.

[11] *An Enquiry Concerning the Principles of Morals,* Appendix 1.

[12] Although I borrow the term "projectivism' from Blackburn—see *Spreading the Word* (New York: Oxford, 1984)—my use of it does not coincide with his. Jackendoff, *op. cit.,* also speaks of a "projected world."

[13] *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin, 1977), pp. 48/9.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
thought with error. There is no property of entitivity in the natural world, says the projectivist. It is a mistake to suppose that some things (or unity schemes) have entitivity while others lack it. It is just a propensity of the mind to interpret the world this way.

This conclusion, if correct, would not be devoid of significance. I am strongly tempted to suppose that I (the present Alvin Goldman stage) shall not be object unified with a tomorrow stage of Margaret Thatcher, but only with a tomorrow stage of Alvin Goldman. The projectivist tells me, in effect, that this is a mistake. Since the shmerson unity scheme has complete ontological parity with the person unity scheme, no invidious distinction can be drawn between my unity with tomorrow's Alvin Goldman versus my unity with tomorrows's Margaret Thatcher. This is, ostensibly, a substantial metaphysical conclusion, one at variance with common sense.

What, then, is the justification for the projectivist's conclusion? There are two arguments for it. First, it is puzzling how there could be any facts in the natural world which would dictate that certain conglomerations of elements have entitivity and other conglomerations lack it. Second, and more centrally, there is an "alternate explanation" argument.

In arguments over God, the theist claims that the best explanation of religious experience and belief is the existence of God. The atheist replies that there is an alternate explanation in terms of psychological and cultural mechanisms. We do not need to postulate God in order to explain religious phenomena, since we have the other, much more parsimonious form of explanation. Like the atheist, the projectivist asserts that we do not need to postulate genuine possession and nonpossession of entitivity in order to explain people's preferences for certain unity schemes (or certain objects) rather than others. Those preferences can be explained much more parsimoniously in terms of innate psychological constraints.

IV.

I hope I have made enough of a case for unity projectivism to make it seem at least worthy of serious discussion, and hence illustrative of a not–altogether–hollow attempt to put results of psychology to metaphysical ends. But the ramifications of the thesis for the problem of personal identity are unsettling. One might well feel that the indicated invidious distinctions should not be obliterated, as projectivism would do. Can we live with this consequence? If not, how should one respond to unity projectivism?

One natural response is found in reactions to moral projectivism: the secondary–quality response. John McDowell[14] answers Mackie's

[14] "Values and Secondary Qualities," in Ted Honderich, ed., *Morality and Objectivity* (London: Routledge & Kegan Paul, 1985).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

projectivism by suggesting that values have the same real, i.e., non-subjective, status that other secondary qualities have, viz., a power to produce certain effects in us. The same maneuver might be extended to the unity problem. Projectivism seems to be a kind of eliminativism. But a secondary–quality view, as McDowell points out, preserves a legitimate ontological niche for the entities in question, indeed, a *not–purely–subjective* ontological niche. For our topic, it would say that the property of entitivity is the ("objective") property of having the power to produce certain effects in human beings. As such, entitivity is secure, even in a scientific image of the world.

Many questions might be raised for the secondary–quality gambit. But suppose for the sake of argument that it succeeds. This might defeat projectivism, but it does not defeat the suggestion that cognitive science can have ramifications for metaphysics, even revisionary metaphysics. This is because the conclusion that entitivity is a secondary quality is already a metaphysical conclusion, plausibly a revisionary metaphysical conclusion. So the kind of connection I have adumbrated between cognitive science and metaphysics would still be exemplified.

ALVIN I. GOLDMAN

University of Arizona

EPISTEMOLOGY IN THE AGE OF NEUROSCIENCE*

IN a recent archaeological dig through several decades of my papers, I unearthed the graduate student stratum. At that level, I discovered a dusty bundle of final examination papers from Oxford, and, in pondering the test questions, I found myself viewing most of them as old curiosities. Among such curiosities for me are the Gettier problem, the nature of sense data, the nature of incorrigible foundations of knowledge, and the constituents of the corpus of a priori knowledge.

Though the examination papers are in fact less than twenty years old, so much has changed, and changed profoundly, in the field, that it may not be an exaggeration to say that we are in the midst of a

* To be presented in an APA symposium on Epistemology and the Philosophy of Mind, December 28, 1987. Alvin I. Goldman will be co–symposiast, and George Bealer will comment; see this JOURNAL, this issue, 537–544 and 553–555, respectively, for their contributions.

Grateful thanks are owed to Paul Churchland, Patricia Kitcher, Stephen P. Stich, Terrence J. Sejnowski, and Warren Dow for discussion and advice. The three figures below are adapted from Paul Churchland's *Matter and Consciousness*, 2nd ed. (Cambridge, Mass.: MIT, forthcoming).

0022-362X/87/8410/0544$01.00 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
paradigm shift. Since one cannot accurately speak for the whole field, I should properly speak only for myself (and Paul Churchland). With that proviso, it seems to me that the general frame of reference within which we might hope to discover how humans learn, understand, and perceive is undergoing a major reconfiguration. Most of the questions which used to preoccupy us as graduate students and whose answers seemed necessary to advancing the general program in epistemology, now look either peripheral or misguided, and the general program itself looks troubled.

I am no longer preoccupied with the nature of absolute foundations, because it does not look as if there are any; or with a priori knowledge, because there probably is not any, or with sense data, because that is a mixed–up way of thinking about sensory processing. It is doubtful that knowledge in general is sentential; rather, representations are typically structures of a quite different sort. Whatever reasoning and information processing turn out to be, formal logic is probably not the model, save perhaps for a small part. Decision theory, confirmation theory, the predicate calculus, etc., beautiful and magnificently clever though they are, do not appear destined to play a central part in the theory of how, in fact, human and other nervous systems solve problems and figure things out. Inductive logic does not exist, and does not show any positive signs in that direction; 'inference–to–the–best–explanation' is a name for a problem, not a theory of how humans accomplish some task.

Formal semantics now looks like a thoroughly misbegotten project which cannot even begin to explain how human language is meaningful. Hilary Putnam[1] has provided grounds for this view, and empirical data from linguistics also seriously undermines the general idea.[2] Formal semantics may be invaluable for certain purposes, but, for generating a theory of meaningfulness in human language, it looks like a dismal failure. Surely there is something bizarre about the idea that a theory of meaning that has nothing whatever to do with human psychology or neurophysiology can explain the meaningfulness of language and how representational structures relate to the world. And what of truth? If representational structures are not sentences (propositions), they are not truth-valuable; if they are to be evaluated, it must be in some other way.[3] Consequently, the very concept of truth appears to be in for major reconsideration.

[1] *Reason, Truth, and History* (New York: Cambridge, 1981).

[2] George Lakoff, *Women, Fire, and Dangerous Things* (Chicago: University Press, 1987).

[3] For a further discussion of the deep problems with truth and rationality, see Stephen P. Stich, *The Fragmentation of Reason* (Cambridge, Mass.: MIT Press, forthcoming).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Thus, knowledge and belief, reference, meaning, and truth, and reasoning, explaining, and learning, are each the focus of eroded confidence in "the grand old paradigm," a framework derived mainly from Logical Empiricism, whose roots, in turn, reach back to Hume, Locke, and Descartes. This is certainly not to say that nothing has been achieved, or that everything worked out from within the confines of the old assumptions is bunk. On the contrary, there are surely many enduring results, although at this stage I find it difficult to know how to tell the enduring from the ephemeral.

Moreover, it is not that there has been a decisive refutation of "the grand old paradigm." Paradigms rarely fall with decisive refutation; rather, they become enfeebled and slowly lose adherents. Confirmed practitioners can always continue, secure in the faith that a new wrinkle may yet satisfy the critics. But many of us sense that working within "the grand old paradigm" is not very rewarding. By contrast, there is considerable promise in a naturalistic approach, which draws upon what is available in psychology and neuroscience to inform our research. There are remarkable new developments in cognitive neurobiology which encourage us to think that a new and encompassing paradigm is emerging. Epistemology conceived in this spirit is what W. V. Quine[4] has called *naturalized* epistemology.

THE BIOLOGICAL PERSPECTIVE

The fundamental epistemological question from Plato onward is this: How is it possible for us to represent reality? How is it that we can represent the external world of objects, of space and time, of motion and color? How do we represent our inner world of thoughts and desires, images and ideas, self and consciousness? Since it is, after all, the nervous system that achieves these things, the fundamental epistemological question can be reformulated thus: *How does the brain work?* Less cryptically and more accurately, the question is: *How, situated in its bodily configuration, within its surrounding physical environment, and within the social context it finds itself, does the brain work?* Answers here will be descriptive first and foremost, but the normative dimension of epistemology enters when we can draw on the descriptive basis to compare and evaluate styles of computation and representation, and determine how to improve upon particular computational and representational strategies. Once we understand what reasoning is, we can begin to figure out what reasoning *well* is.

This is a good time to be naturalizing epistemology, and three factors in particular indicate this. First, technological developments

[4] "Epistemology Naturalized," *Ontological Relativity and Other Essays* (New York: Columbia, 1969), pp. 69–90.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

in research in the neurosciences during the past twenty years have been spectacular, and a truly impressive amount is now known about the microstructure and micro-organization of nervous systems. For example, we are learning in detail about the pathways for particular neuron types in the visual system, about the physiological properties of different types of neurons, and about the different tasks performed by distinct neural populations. Techniques for addressing nervous systems at a variety of levels of organization have revealed anatomical and physiological data that suggest that the time is ripe for genuine theorizing about how macro effects are the outcome of neuronal properties.[5]

Second, the advent of cheap computing makes it possible to simulate neural networks and, hence, to investigate computational properties at the *circuit* level. Since we do not have physiological techniques to address this level in actual nervous systems, the computer simulation of biological circuits is a crucial adjunct to available neuroscientific techniques. Cheap, fast computing is essential for simulation, because nervous systems have a parallel architecture, and huge numbers of computational events are going on simultaneously in the network. Nevertheless, even new computing technology such as the current-generation connection machine[6] is still far inferior to the human brain in computing capacity. The communications bandwidth of the connection machine is about 10^{10} bits per second, which is extraordinary. But the *average* communications bandwidth used by the human brain is about (10^{11} neurons)(5×10^3 connections/neurons)(2 bits/connection/sec), which means that the brain processes on average 10^{15} bits/second.[7]

Third, studies of animal behavior in ethology, psychology, linguistics, and clinical neurology have become increasingly sophisticated and have yielded crucial and often surprising data concerning the capacities of nervous systems. Within clinical neurology, one of the most important discoveries concerns the multiplicity of memory systems in humans and in monkeys.[8] Within psychology, the discoveries that virtually all categories show prototype effects[9] and that images

[5] See my *Neurophilosophy* (Cambridge, Mass.: MIT Press, 1986).

[6] W. Daniel Hillis, *The Connection Machine* (Cambridge, Mass.: MIT Press, 1985).

[7] Terrence J. Sejnowski, review of *The Connection Machine, Journal of Mathematical Psychology*, XXXI (1987): in press.

[8] This is thoroughly and clearly discussed by Larry Squire in *Memory and Brain* (New York: Oxford, 1987).

[9] Eleanor Rosch, "Human Categorization," in N. Warren, ed., *Studies in Cross-cultural Psychology* (London: Academic Press, 1977).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

are prevalent in information processing[10] stand out as particularly important.

Adopting the biological perspective, we find that a number of points have special significance. Not that these points are generally unknown, but from a biological perspective, they acquire distinctive salience, for they shape the way we think about the problems of vision, learning, spatial representation, sensorimotor control, and so forth.

The most fundamental point is that the human brain is a product of evolution. This is worth reflecting on for three reasons: (1) In many important respects, the human brain is very similar in structural components and organization to other primate brains, and not very different from other mammalian brains. Cortical organization, for example, seems generally shared among mammals. Our brains share some basic properties with even the simplest nervous system. Like ours, mammalian nervous systems are networks of connected units; these units, neurons, function basically in the same way in all nervous systems; we share the same neurochemicals. There are bound to be some differences, at least because the human communication capacity appears to be distinctive, but this is probably owed not to a special "wonder lobe" but to subtle, fine–grained wiring differences. The anatomical and physiological similarities are important, because they mean that we should expect that how *we* learn, remember, see, hear, and respect, is not fundamentally different from how other organisms do those things.

(2) Cognition is not neatly detachable from the organism's ecological niche, way of life, and bodily structure. Nervous systems are not general–purpose computers. They have evolved to accomplish a certain range of tasks, and the architecture supports those tasks. There is a fatal tendency to think of the brain as essentially in the fact–finding business—as a device whose primary function is to acquire propositional knowledge. At its best, supposedly, it discovers truth–for–its–own–sake. From a biological perspective, however, this does not make much sense.

Looked at from an evolutionary point of view, the principal function of nervous systems is to enable the organism to move appropriately.[11] Boiled down to essentials, a nervous system enables the organism to succeed in the four F's: feeding, fleeing, fighting, and reproducing. The principal chore of nervous systems is to get the body

[10] Stephen Kosslyn, *Image and Mind* (Cambridge, Mass.: Harvard, 1980).
[11] J. L. McClelland and D. E. Rumelhart, eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Cambridge, Mass.: MIT Press, 1986).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

parts where they should be in order that the organism may survive. Insofar as representations serve that function, representations are a good thing. Getting things right in space and time, therefore, is a crucially important factor for nervous systems, and there is often considerable evolutionary pressure deriving from considerations of speed. Improvements in sensorimotor control confer an evolutionary advantage: a fancier style of representing is advantageous *so long as it is geared to the organism's way of life and enhances the organism's chances of survival.* Truth, whatever that is, definitely takes the hindmost.

(3) We cannot expect engineering perfection in the products of evolution. Improvements in nervous systems are not built by starting from scratch, even though that might yield the best design. They are modifications of structures and patterns that already exist. Evolution has to achieve improved organization by managing with what is there, not by going back to redesign the basics. As we consider research strategy, this is an important consideration. It means that, if we approach the problems of nervous–system function as strictly engineering problems, setting our goals in terms of how a task (for example, stereopsis) could *in principle* be done, we may find a cunning engineering solution which is nothing like the solution the brain has found. In framing hypotheses concerning brain function, it will be essential to consider constraints not only at the behavioral level, but also at the neurobiological level. Unless we go *into* the black box, we run the considerable risk of wasting our time exploring remote, if temporarily fashionable, areas of computational space.

WHY CONNECTIONISM IS IMPORTANT

If representational structures are not sentence–like, what are they? If computation is not logic–like, what is it like? And how can we find out? If formal semantics is the wrong approach to explaining meaningfulness, what would work? These seem to me to be the central questions, or at least some of the questions relevant to epistemology. Connectionism (also known as Parallel Distributed Processing, or PDP) is important, because it constitutes the beginnings of a genuine, systematic alternative to the "grand old paradigm." It appears to have the resources to provide neurobiologically plausible answers to these central questions (see McClelland and Rumelhart, *op. cit.*).

Connectionist models illustrate what representations might really be, if not sentence–like, and what neurobiological and psychological computation might really be, if not logic–like. They free us from the conviction that the sentence/logic model is inevitable. Moreover, the design of the models is inspired and informed by neurobiology; so they are more biologically realistic than sentence/logic models.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

A connectionist model is characterized by three architectural elements: (1) processing units, (2) connections between processing units, and (3) weights, which are differential strengths of connection between processing units. The processing units, like neurons, communicate with each other by signals (such as firing rate) which are numerical rather than symbolic. In typical models, the units are arranged in layers: an input layer, an output layer, and, intervening between, the layer of so–called "hidden units" which are fully connected to the input and output layers (see Figure 1). The processing units sum the inputs from the connections with other processing units, each input weighted by the strength of connection. The output of each processing unit is a real number that is a nonlinear function of the linearly summed inputs (see Figure 2). The output is small when the inputs are below threshold, and increase rapidly as the total input becomes more positive. To a first approximation, the activity level can be considered the summed postsynaptic potentials in a neuron, and the output can be considered its firing rate.

What is remarkable about such systems is that they can be trained to learn highly complex input–output functions. This learning is not merely a look–up table achievement, because the small number of hidden units makes that impossible. Moreover, the system can gener-

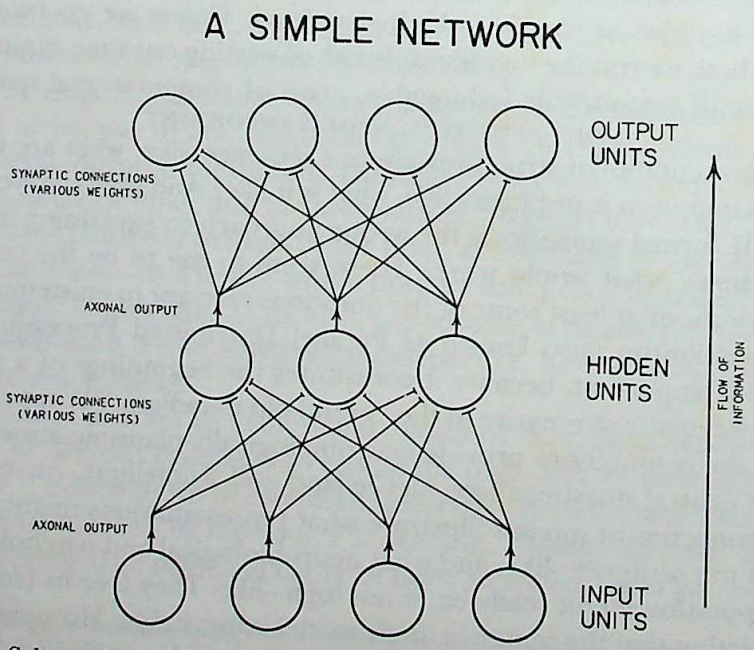


Fig. 1: Schematic model of a three-layered network. Each layer is fully connected to the layer adjacent to it.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

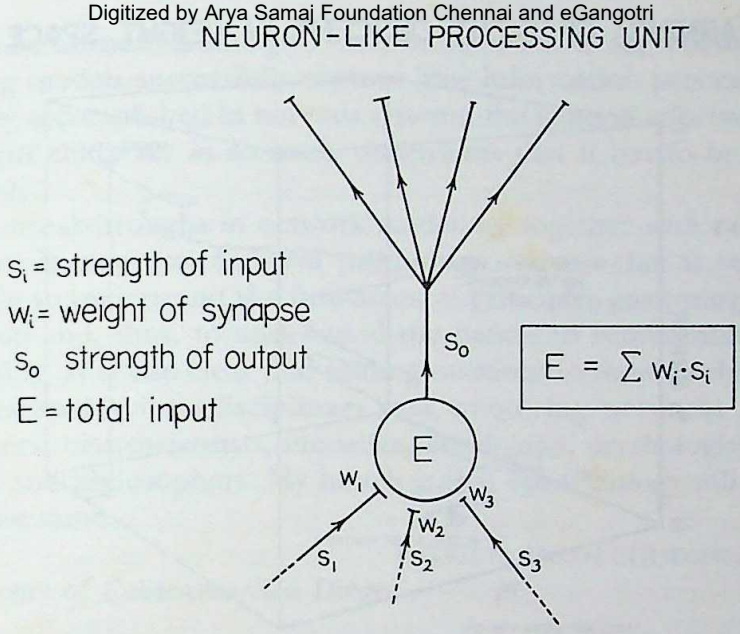


Fig. 2. Schematic model of a processing unit receiving inputs from other processing units.

alize to new cases, and it can make sensible approximations in early training or when there is no correct answer.

NETtalk[12] is perhaps the most spectacular example of a connectionist network. It learns to convert written text to speech. The network does not have any initial or built–in organization for processing the input, or, more exactly, for mapping letters onto sounds. All such organization emerges during the training period. The values of the weights are determined by using the learning algorithm called "back–propagation of error."[13] The strategy exploits the calculated error between the *actual* values of the processing units in the output layer and the *desired* values, which are provided by a training signal. The error signal is propagated from the output layer backward to the input layer and is used to adjust each weight in the network. The network learns as the weights are changed to minimize the mean squared error over the training set of words. Thus, the system can be characterized as following a path in weight space until it finds an error minimum (see Figure 3). The important point to be illustrated,

[12] Terrence J. Sejnowski and Charles R. Rosenberg, "Parallel Networks that Learn to Pronounce English Text," *Complex Systems*, I (1987): 145–168.

[13] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representations by Error Propagation," in McClelland and Rumelhart, eds., *op. cit.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

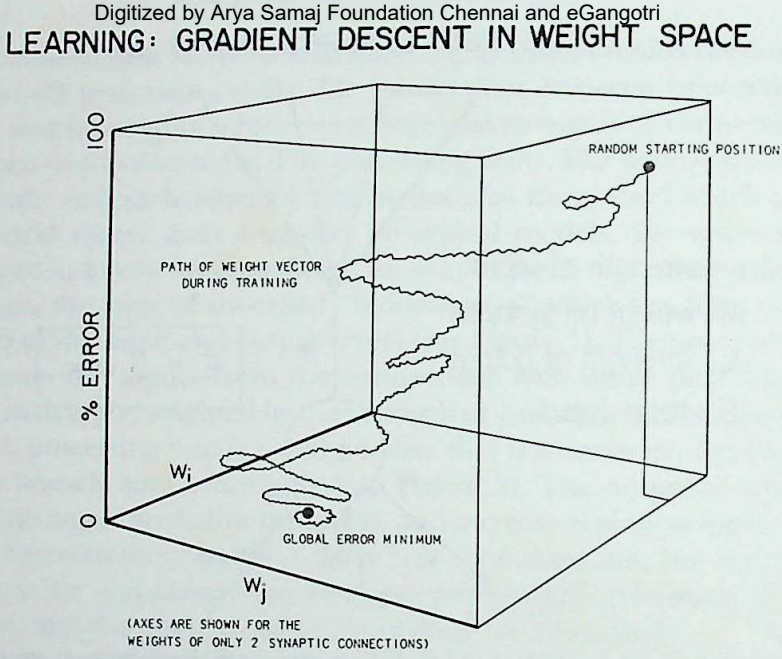


Fig. 3. Schematic drawing of a path followed in weight space as the network finds a global minimum.

therefore, is this: the network processes information by nonlinear dynamics, not by symbol manipulation and rule following. It learns by gradient descent in a complex interactive system, not by generating new rules. Representations in the hidden units turn out to be not symbols, but patterns of activation (see Sejnowski and Rosenberg, *op. cit.*).

The general strategy in connectionism is to model information processing in terms of the trajectory of a complex, nonlinear dynamical system in a very high dimensional space. This structure does not resemble sentences arrayed in logical sequences, but it is potentially rich enough and complex enough to yield behavior that can support semantic and logical relationships.

The connectionist approach is still very new, and many questions remain unanswered, but progress so far is impressive. Some models address specific neurobiological problems and are highly constrained by neurobiological data. Others address a higher level of organization. At this stage, the important thing is that a model suggest experiments, preferably at the neurobiological level, where the results will suggest further modifications in the model. Additionally, it is now imperative that additional information-processing algo-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

rithms be devised. Although it may be too soon to say whether any existing models successfully capture how information processing is actually accomplished in nervous systems, the general approach has the right character in so many dimensions that it has to be taken seriously.

The breakthroughs in network modeling, together with new discoveries in neuroscience and psychology, suggest that it really is possible to understand the fundamental principles governing brain function and, thus, to understand the nature of representing and reasoning. It is also clear that finding solutions to these problems is an inescapably interdisciplinary task, requiring networks of researchers: neuroscientists, modelers, ethologists, psychologists, linguists, and philosophers. My hunch is that epistemology will never look the same.

PATRICIA SMITH CHURCHLAND
University of California/San Diego

THE BOUNDARY BETWEEN PHILOSOPHY AND COGNITIVE SCIENCE*

The main issue before us concerns the boundary between philosophy and empirical science. Patricia Smith Churchland and Alvin I. Goldman advocate an aggressive empiricist philosophy according to which many fundamental philosophical questions lie within the province of empirical science—specifically, cognitive psychology, physiology, and perhaps experimentally based computer science. The primary problem (not to say the only problem) with this aggressive empiricism concerns modality. Like the answers to questions in pure mathematics, the answers to basic philosophical questions are *necessary if true*. For example, if justified true belief is not knowledge, then necessarily justified true belief is not knowledge. Even if all and only cases of justified true belief were, in fact, cases of knowledge, that would not show that justified true belief is knowledge; for the mere possibility of a case of justified true belief that is not knowledge suffices to show that justified true belief is not knowledge.

* Abstract of a paper to be presented in an APA symposium on Epistemology and Philosophy of Mind, December 28, 1987, commenting on papers by Alvin I. Goldman and Patricia Smith Churchland, this JOURNAL, this issue, 537–544 and 544–553, respectively.

0022-362X/87/8410/0553$00.50 © 1987 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

In the history of philosophy, knowledge of necessities has been thought to lie outside the province of empirical science in the following sense: if it is possible to know something to be necessary, it is possible, in principle, to know it to be necessary without the aid of empirical science; when it comes to knowledge of necessities, empirical science is never essential. In recent years, however, scientific essentialists (Saul Kripke, Hilary Putnam, and their followers) have given persuasive arguments that certain things can be known to be necessary only with the assistance of empirical science (for example, that water is H_2O, that gold is the element with atomic number seventy-nine, and so forth). Is knowledge of philosophical necessities like this, too? Can scientific essentialism be generalized from philosophically uninteresting, naturalistic issues (such as the nature of water or gold) to basic philosophical issues? I will argue that this global generalization is doomed. The reason is that the only satisfactory account of scientifically grounded knowledge of naturalistic necessities (water, heat, etc.) presupposes a circumscribed form of rationalism. This circumscribed rationalism entails the reliability (perhaps after theoretical systematization or dialectical critique) of intuitions concerning the applicability of category and content concepts to hypothetical cases that are characterized exclusively in terms of category and content concepts. Most of the basic questions of philosophy, however, are put exclusively in terms of category and content concepts. These considerations imply a thesis of the *autonomy of philosophy:* for most philosophical questions, if it is possible to know the answers to these questions, it is, in principle, possible to know the answers to them without the aid of empirical science.

Accordingly, when Churchland seeks to learn what "knowledge and belief, reference, meaning, and truth, and reasoning, explaining, and learning" *are,* empirical sciences (empirical psychology, physiology, and computer modeling) are incidental, at best. Our concepts of these items are basic category and content concepts, so necessities involving these concepts are, if knowable at all, knowable without the aid of empirical science.

Similarly, when Goldman tries to understand unity and objectivity, empirical psychology can be incidental, at best; for the concepts of unity and objectivity are category concepts. As such, their analysis (what unity and objectivity are supposed to be) lies outside the province of empirical science. (I believe that the right analysis must be given in terms of a special kind of purely logical theory.) But how is one to decide which items are objective unities, for example, whether persons are and Goldman's "shmersons" (if they exist) are not? We

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

must certainly consult our best overall theory, which will be partly empirical. Does this show that the question is empirical? Not at all. For independent of any particular empirical considerations, there are transcendental considerations showing that any acceptable comprehensive theory of the world must confer on certain items (for example, the beings who actually construct such theories) special objective status. Persons (but not shmersons) are like this.

GEORGE BEALER

Reed College

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

LIBERTIES, WRONGS, AND REPRESENTATION*

IT is no small irony that liberal democracies leave some groups of citizens underrepresented. An important reason for preferring liberal representative government is that, without it, some compelling citizen interests can be ignored, and so liberal democracies in which compelling interests can still be ignored seem to fail by their own standard. When we think of philosophers who have taken up this problem, we might well think first of Alexis de Tocqueville and John Stuart Mill. We would not think first of Hegel. In my view that is our misfortune. In what follows, I offer an interpretation and a defense of the Hegelian view.

I.

The best expression of Hegel's position appeared in his *Rechtsphilosophie* lectures of 1824/5:

> Organizing from top downwards has been done for some time, and this organizing has been the main effort. Great efforts have been made to organize chambers and estates, but at the bottom of the whole the masses were carelessly left more or less unorganized, and yet it is of the highest importance that this become organic, for only in this way is it a force and an authority, otherwise it is a heap, a crowd of divided atoms, a crowd which has no rational authority, a blind force.
>
> Legitimate authority is present only in cases where the particular spheres are organized. This side—the particular—is among the most important, and in France the need is becoming ever greater; indeed one still cannot save ministerial direction from bureaucracy. Thus the rest of the constitution hangs in the wind, not having its feet rooted in what is organic, legitimate, and independent.[1]

Three reasons are given here for extending political organization beneath the parliamentary level to the masses below *(das Untere, das Massenhafte des Ganzen)*. One reason is that, if they are not so organized, then they have no authority *(Gewalt, vernüntftige Gewalt)*. Those who are to be organized ought to have some force and authority, which only organization can provide.[2] *Nur so ist es Macht.*

* To be presented in an APA symposium of the same title, December 29, 1987. Martin J. De Nys will comment; see this JOURNAL, this issue, 565/6.

I am grateful to Frank Cunningham and Calvin Normore for their comments on earlier versions of this paper.

[1] G. W. F. Hegel, *Vorlesungen über Rechtsphilosophie 1818–1831*, 6 vols., Karl-Heinz Ilting, ed. (Stuttgart: Frommann-Holzboog, 1973), vol. 4, pp. 692/3.

[2] As I will show, the reason is this: they have legitimate grievances (concerning poverty, among other things) that it is not safe to entrust solely to parliamentary representatives. Thus, they must retain some separate authority, exercised through nonparliamentary channels.

0022-362X/87/8410/0556$00.90 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

At first glance, this seems inconsistent with Hegel's second reason for organizing political masses: without such organization, they are a blind force *(eine blinde Macht)*. The inconsistency between these two phrases is merely verbal. Any group that can legitimately be called "oppressed" is certain to have legitimate grievances and is likely to be powerless against their cause, and such groups are not unknown to explode violently against the substance and symbols of property and the state. Hegel had this sort of mob action in mind when he called political masses "blind forces." One political pamphlet he wrote threatened that calls for reform, if unheeded, would be followed by "a more frightful outburst, in which the need for reform joins hands with revenge, and the mob, ever deceived and oppressed, visits dishonesty with punishment."[3] It is, in part, because a political mass lacks power *(Macht)* that it becomes a blind force *(eine blinde Macht)*. Without political organization members of a political mass are powerless, and, if they were not powerless, they would not explode as a blind force.

Some interpreters claim that Hegel's *only* reason for proposing that political masses be organized was that they would *otherwise* continue to act as a subversive, destabilizing force threatening to undermine the modern state. This is quite incorrect. Hegel's notes for the English Reform Bill essay of 1831 express his objection to rioting and his sympathy with the rioters in two consecutive sentences: (1) "If . . . the government, out of some such motive, were to revolt against the law and let plunder, would this also be excused?"[4] (2) "One must not forget how vexed [the people are]," he went on, "they are [acting] wrong[ly] against the law, but [they are acting] on [a] justified principle" *(ibid.)*. Thus, failure to provide political masses with organization and power is responsible, in Hegel's view, for two distinct evils. Without organization and power, they will not infrequently place the security of property and the state at risk; at the same time, their own legitimate grievances will go unresolved.

The third reason we find in this passage for organizing and empowering political masses is that the legitimacy of any modern state that does not do so is in jeopardy. When Hegel says, "Legitimate authority is present only in cases where the particular spheres are

[3] "On the Recent Domestic Affairs of Württemberg, Especially on the Inadequacy of the Municipal Constitution," *Hegel's Political Writings*, Z. A. Pelczynski, ed., T. M. Knox, trans. (New York: Oxford, 1964), p. 243. A similar threat was delivered at the outset of Hegel's 1831 essay on the English Reform Bill, *ibid.*, p. 258.

[4] Hegel, "Zwei Entwürfe zur Reformbill-Schrift," *Werke*, 20 vols. (Frankfurt: Suhrkamp, 1969–71), vol. 11, p. 553.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

organized," I take this to mean that legitimate authority is present only in those *states* in which the particular spheres are organized. In the absence of such organization, the "rest of the constitution hangs in the wind, not having its feet rooted in what is organic, legitimate, and independent."

It is important to distinguish the sort of independence that Hegel prescribes here for legitimate interest groups from the sort of independence that he criticizes elsewhere. "In feudal times," he complained, ". . . the particular functions and powers of the state and civil society were arranged . . . into independent Corporations and societies, so that the state as a whole was rather an aggregate than an organism."[5] To capture the difference between those two sorts of independence, one of which Hegel recommends and the other of which he considers pathological,[6] I want to rename the more valuable form of independence *autonomy*. Both 'independence' and 'autonomy' can be defined in terms of particular interests. If we define any subset of the population by a condition that all and only its members share, then we may find that, *because* they share this condition, there are some outcomes that it is rational for all of them to prefer, but not rational for anyone else to prefer. These outcomes comprise the *particular interests* of that group.

Particular interests conflict when it is in the interests of one group to bring about an outcome that it is in the interests of another group to prevent. Call any group that can make its interests prevail in case of conflict *independent*, and any a group that cannot do so *dependent*. Since it cannot make its interests prevail, the dependent group is forced to revise its conception of where its interests lie. If neither group can make its interests prevail, then each may have to reconsider its interests with a view to compromise. This was the outcome Hegel preferred in cases of conflict between otherwise legitimate interests—"the reconciling affirmation, the 'yes', with which both egos desist from their existence in opposition."[7] In this case, though neither group can achieve its interest independently, we might still say that both are *autonomous*. To have autonomy is to have not the power to impose what is in one's own interests, but rather the power to force a compromise, in case of conflict.

An interest group can be autonomous in the pertinent sense only if it is capable of reconsidering its interests, and, for this capacity,

[5] *Philosophy of Right*, T. M. Knox, ed. and trans. (New York: Oxford, 1952), §278 Zusatz, p. 180.

[6] *Ibid.*, "isolation and independence" of the parts of a state "spell disease."

[7] *Phenomenology of Mind*, J. B. Baillie, trans. (New York: Harper & Row, 1967), p. 679.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
some capacity for deliberation is necessary. Autonomy in this sense is not achievable by a political mass, which lacks the internal organization necessary for group deliberation. Consequently, organization is a necessary condition for autonomy. Any group that needs autonomy needs to be organized.

But does anyone really need political autonomy? As I interpret Hegel, he said that all particular interest groups need political autonomy. His reasons can be distilled from his analysis of the French Revolution. The main question arising in the aftermath of the revolution, he thought, was this: The republican state must enact what is in everyone's interest; but should these acts of the nation also serve particular interests? The Terror of the early 1790s was, in Hegel's view, the practical outcome of answering this question in the negative. Since that answer is self–defeating, shall we answer affirmatively, that the means by which the nation serves the interests of all *may* also serve particular interests? The practical consequence of this alternative is parliamentary factionalism.

If we allow special interests to benefit from the state operations that execute the universal interest, we can be sure that these benefits will be seized. We can also expect a coalition of interest groups to form with the following goals in mind: (a) to hold state power, (b) to ensure that the state apparatus works in ways that coincidentally serve its own particular interests, and (c) to maximize these benefits. Maximization of benefits requires keeping the coalition as small as possible—without sacrificing its hold on power. Even if coalitions shift and power changes hands, it remains possible for some special interest to be excluded from every ruling coalition. It matters not at all whether this is a compelling interest—one that ought not to be ignored. Parliamentary factionalism allows any particular interest to be ignored—unless, of course, that particular interest group has the power to force a compromise.

Consider any interest group that has, among its particular interests, some that are politically compelling—I mean interests that no legitimate state can ignore. Because they are prone to factionalism, parliamentary states cannot ensure that these interests will not be ignored, unless they provide such a group with the power to protect its compelling interests by forcing a compromise upon the ruling faction. If there are any such groups, then a parliamentary state that does not give them political autonomy places its own legitimacy in jeopardy. "Thus the rest of the constitution hangs in the wind."

II.

Here we encounter a dilemma. How shall we put political autonomy into effect? Either we contrive to give political autonomy only to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

those groups which have compelling interests—and which run significant risks that these interests will otherwise be ignored—or else we arrange to give political autonomy to all particular interest groups. In the first option, a principle of equal liberty seems to be violated, since there is a form of political representation from which part of the population is excluded. But the second option is likely to be ineffective in protecting compelling interests *other than* an interest in equal liberty. Either way, some compelling interests seem to be left unprotected.

It is in order to avoid the first horn of the dilemma that one is most likely to entertain the second, the pluralist option—that every particular interest group should have the means to protect its particular compelling interests by forcing a compromise if necessary. What blocks the pluralist option primarily is the nature of these means themselves. Of the compromise–forcing devices that might be used, there are not likely to be any which can be used to protect everyone. The instruments I have in mind include such things as limited veto powers and powers to nominate, confirm, or recall appointed or elected officials. Powers like these must be tightly circumscribed, so as to prevent their abuse. In particular, we should want to prevent anyone from using them to advance or protect additional interests beyond those compelling interests which the state has a vital interest to protect.

Now, some groups share particular interests of which none are compelling. For instance, it strains credibility to think that the legitimacy of any state should depend on the protection it affords to aficionados of cockfighting. A more serious case is that of the *elite group,* by which I shall mean a group whose particular interests are so well entrenched by other means that they do not need the further protection of political autonomy. Since social relations overlap, some members of an elite may also belong to nonelite groups that do require political autonomy. But suppose that some do not; suppose, as is likely, that there is at least one member of an elite group who belongs to no such nonelites. Equal liberty seems to require that the elite should have the protection of compromise-forcing powers, so that *every individual* somehow enjoys this protection. But this is impossible, since the instruments of political autonomy must accord with the group's need, and the elite group's need is by definition nil. Pluralism is impossible unless every elite member gains access to the means of political autonomy by belonging to a protected nonelite. But this is unlikely.

Hegel's corporatism attempts—unsuccessfully—to solve this problem. Hegel wanted to reconstruct the medieval guild and town

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
organizations as vehicles through which political rights were exercised.[8] Hegel was a modern Cleisthenes, but, instead of giving everyone a tribe, he wished to give everyone a guild.[9] If elite groups comprised minorities within each corporation, then every individual could be represented by a group with political autonomy—the corporation—without requiring the instruments of autonomy to be given to any group that did not need them. For instance, in a corporation including workers, owners, and managers in a single industry, it is the workers who, in virtue of their economic insecurity, have the compelling interests that need protection. On what I presume to be Hegel's analysis, they share with their employers an interest in the economic health of their industry. The worker majority can use the corporation's compromise–forcing powers to block major threats to the health of the industry so as to protect itself from unemployment. Since this incidentally benefits the elite groups of managers and owners, no member of the corporation is deprived of protection by compromise–forcing powers; yet these powers cannot be exploited for the particular interests of the elite, either.

There are, however, two reasons why this solution may not succeed after all. First, it may be true that the interests of labor and capital are in the long run irreconcilable. If so, we can expect the elite within any of Hegel's industry–wide corporations to form coalitions with middle strata, possibly to the disadvantage of those members who are most in need of protection. Unless the interests of the corporation's *members* are nearly all the same, the play of factions can erupt, jeopardizing the compelling interests that were to be protected. Secondly, not all groups who need the protection of political autonomy would form majorities in industrial or municipal organizations. If, unlike Hegel, we recognize women and national minorities as groups that need such protection, then we must recognize that Hegel's corporations will not provide it.

Therefore, we cannot avoid the other horn of the dilemma and, with it, the principle that everyone should be equal in basic liberties. On the face of it, any of our schemes to give compromise–forcing powers exclusively to deserving interest groups offends against this principle. What is especially interesting about the Hegelian view,

[8] Hegel followed legal tradition by including town as well as craft organizations among corporations. See *Philosophy of Right*, §288, p. 189; addition to §252, p. 153; §251, p. 152.

[9] The nobility had no corporation, but they were to have the entire upper chamber to themselves; civil servants were given no corporation, but Hegel may have intended them to join municipal corporations.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

though, is its suggestion that equal–liberty considerations may recommend special representation rather than forbid it.

I began this paper speaking of groups with compelling interests, and by these I meant interests of such a kind that a state that could not protect them risked losing its legitimacy. Now I must say more about who these groups are. In some important passages, Hegel says that the state must especially protect groups who are *socially wronged.* He believed that the poor in general and the unemployed in particular were victims of a "wrong done to one class by another."[10] Since rational authority does not allow wrongs to continue, the legitimacy of any state that does allow them to continue is questionable.

Hegel's conception of social wrongs rests on an analogy to civil wrongs, as in the law of torts. In Hegel's book, torts as well as crimes are injurious in two respects. First, they are damaging to one's property, to one's body, or to one's other assets—in short, to things which "exist externally" (*ibid.*, §98, p. 69). Further injury is done to what Hegel calls the "*implicit* will," the will *an sich* (*ibid.*, §99, p. 69). Here this means that there is some *as yet unexecuted will* that is injured by torts and crimes. He identifies this will as "the right [*das Recht*] or law implicit" (*ibid.*). If pauperization wrongs a social class, then, by analogy, it must not only injure them physically, but it must also injure this mysterious will which is said to be "implicit." To make a long story short, this was, in Hegel's view, the will that came to be executed in the course of history, and, according to his own summaries of the course of history, the object of this will is the equalization of liberty. In the Orient, "Only One is free"; in ancient Greece and Rome, "only Some are free"; in modern Europe, "man as such is free."[11] The discovery that a particular form of cooperation does not promote equal freedom costs the regime its legitimacy, and delegitimization leads ultimately to its demise, on the Hegelian theory.

Hegel had made just such a discovery in connection with the effects of poverty in a market economy. The unemployed are unfree by a standard of self–reliance. Hegel notes not only their "distress" but their "dependency" (*Philosophy of Right,* §243, pp. 149/50), which offends against the "self-respect which makes a man insist on maintaining himself by his own work and effort" (§244, p. 150). As well, dependency "would violate the principle of civil society and the

[10] *Philosophy of Right,* §278. Cf. Ilting ed., vol. 4, p. 609.
[11] *Lectures on the Philosophy of World History, Introduction: Reason in History,* Johannes Hoffmeister, ed., H. B. Nisbet, trans. (New York: Cambridge, 1975), p. 130.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
feeling of individual independence and self–respect in its individual members" (§245, p. 150). Poverty and unemployment can indeed injure their victims physically, but in addition they commit an intangible injury. Their effect within modern society, and among its members, is unequal freedom.

To the charge that special representation and political autonomy for the unemployed infringe upon the equal liberty of others, Hegel is able to respond that it is a necessary countermeasure against the condition of unequal liberty which unemployment and poverty bring about. But, without further elaboration, this is not an especially cogent response. For, as the libertarian will quickly respond, one alleged infringement on liberty has merely been replaced by another, which is no less pernicious. What is needed to defend Hegel's insight is a device for showing that the liberty that one side gains—here through being protected against unemployment—is *greater than* the liberty that the other side loses—here by not receiving the political autonomy that protects the former.

By saying that someone is freer than someone else, I mean that there is something which could, if desired, be done by the one but not the other. In what respects, then, should people be equally free? What needs to be held equal is not liberty generally, but rather *defensible basic liberty,* which I define as follows. To have a basic liberty is to be free to perform a basic action. Basic actions are those which have the following property: by doing a basic act of a certain type, you enable yourself to do many further things, from which you are disabled if you perform no acts of that type. Some of the broadest types that basic acts exemplify are: (a) procuring goods to meet one's needs, (b) learning, (c) caring for one's health, and (d) doing something that is necessary to achieve self–respect. If one is restricted from performing actions of any one of these types, then there are further consequences, not only for one's welfare, but also for one's freedom. Restriction of basic actions has a multiplier effect on the balance of one's freedom, since it is quite unlikely that one's performance of basic actions can be impeded without also limiting what *else* one is free to do. The effect, then, of a disadvantage in basic liberties is that, whatever one's desires may be, one's chances of satisfaction are diminished.

Now, I want to say something brief and admittedly incomplete about the *defensibility* of basic actions. Basic actions can be divided into three groups. Through some, we develop our useful powers; through others, we arrange to deploy them and to draw some of the benefits. A third group includes things we must do for self–respect. There are, of course, many ways in which to develop, deploy, and
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

benefit from one's useful powers, not all of which are defensible. The case is even clearer for actions preserving self–respect. No self–respecting mass murderer lets his quarry go unharmed; so homicide is by this criterion necessary for someone's self–respect and, thus, a basic action; from which it follows that murder laws restrict basic liberties. My response is that, although what is not a basic action for some people may indeed be basic for others, still, by recognizing that some pattern of behavior is basic for anyone, we do not imply our consent or approval. It is one thing for behavior to *be* basic for someone and another thing for it to be *defensibly* basic. The principle at stake does not require that everyone's basic liberties be equal. It requires that everyone's defensible basic liberties be equal. That is, all persons must have equivalent opportunities to develop, deploy, and benefit from their useful powers and to preserve self–respect *by defensible means.*

We can easly recognize that it *is* basic to most people, by reason of self–respect, to be economically self–reliant, as Hegel claims. Is this defensible? On the face of it, it must be; for it harms no one else materially, nor can it plausibly be made out to offend against anyone else's dignity. It would not be defensible to link self–respect with a type of act if such acts prevented others from achieving equal self–respect. Thus, it is not defensible to link self–respect either with terrorism or even with *winning,* since both would achieve self–respect by costing others theirs. For the same reason, it is indefensible to require large–scale property acquisition for one's self–respect. But, by achieving economic self–reliance, on the other hand, one makes no one else unable to do so. Thus, we can not only recognize a linkage between self–reliance and self–respect, but we can defend it, too, on grounds of welfare and on grounds of universalizability.

Acts that establish and maintain self–reliance are basic, and defensibly so. To be free to be self–reliant, then, is a basic liberty. By comparison, there is no basic action that is impaired by a regime which withholds political autonomy from interest groups which do not need it. So if, by providing political autonomy to the economically insecure elements of the population, we provide them with a basic liberty, while depriving the others of nothing so basic, then this scheme of differential representation equalizes basic liberties that would otherwise remain unequal. Equal basic liberty may require unequal distribution of power.

JAY DRYDYK

Erindale College
University of Toronto/Mississauga

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

POLITICAL REPRESENTATION AND ECONOMIC LIBERTY*

The state, on Hegel's understanding, is the concrete actuality of freedom. It prescribes laws that specify what is fundamentally and comprehensively right, thereby pursuing universal ends. It is also a domain in which citizens, whose actions foster the universal ends of the state, satisfy particular interests as well. And the state is a political community wherein different groups of citizens, distinguished by diverse particular interests which do conflict, find that their interactions can reciprocally serve to satisfy those particular interests as well as universal political ends.

Jay Drydyk is quite correct in his interpretation of Hegel's idea of the modern state in three important respects. First, the state, in pursuing universal ends, must recognize the right of the particular. Second, particular interest groups require the political organization that allows them to pursue their own interests while simultaneously acknowledging the universal ends of the state. Third, the form of political organization that particular interest groups require must institute the type of independence that Drydyk calls "autonomy." Hegel does not speak of autonomy and compromise in just the way Drydyk recommends. But this recommendation enables one to conceive of political interactions among distinct interest groups which, because they resolve in compromise, do reciprocally serve the interests of the groups in question, enabling those groups to recognize their interactions as interdependent, as serving both particular interests and universal ends. This recommendation, then, is a legitimate development of the Hegelian position.

But how is the state to bring about political autonomy? More specifically, can the state allocate to those who risk poverty and unemployment, to those whose defensible basic liberties of self-reliance and self–respect are at risk, special representation and power, for the sake of bringing about a certain kind of equal liberty? Hegel's theory of the state as a concrete domain of freedom may diffuse abstract libertarian objections to this proposal. But Hegel's own views about the relation of the state to civil society seem to pose objections to this proposal.

* Abstract of a paper to be presented in an APA symposium on Liberties, Wrongs, and Representation, December 29, 1987, commenting on Jay Drydyk's paper of the same title, this JOURNAL, this issue, 556–564.

0022-362X/87/8410/0565$00.50 © 1987 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The state, Hegel believes, both grounds and presupposes civil society. The latter is the case because civil society, a system of actions and interactions wherein self–interest is pursued in a market economy, is the social domain in which the right of the particular asserts itself. A market economy is also, Hegel argues, an economy in which unemployment and poverty are systemically unavoidable. If so, then political power, used by the potentially unemployed to ward off the possibility of unemployment and poverty, would be ineffective unless it were aimed at transforming the market economy in some fundamental ways. But this would be an attack on the market economy which the state cannot permit, Hegel believes, since the state has that economy as its own presupposition.

Hegel does, arguably, show that the legitimacy of the modern state follows from the fact that it promotes equality in something like defensible basic liberties. But his own arguments also show that the defensible basic liberties of those who risk unemployment and poverty cannot be ensured through political processes alone. Hegel's own arguments suggest, in spite of his interpretation of those arguments, that the state requires transformations in its economic presupposition if it is to find legitimacy in being the concrete actuality of freedom. The contemporary interpreter of Hegel is challenged to contrive an understanding of the economic domain of social life which allows for the assertion of the right of the particular and which does not systematically lead to poverty and unemployment. The problem of equality in defensible basic liberties is not only political, but also more primitively an economic problem.

MARTIN J. DE NYS

George Mason University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

HOW GOD CAUSES MOTION: DESCARTES, DIVINE SUSTENANCE, AND OCCASIONALISM*

IN his *Traité de l'esprit de l'homme* (1664),[1] Louis de la Forge, one of Descartes's early followers, wrote:

> I hold that there is no creature, spiritual or corporeal, that can change [the position of a body] or that of any of its parts in the second instant of its creation if the creator does not do it himself, since it is he who had produced this part of matter in place A. For example, not only is it necessary that he continue to produce it if he wants it to continue to exist, but also, since he cannot create it everywhere, nor can he create it outside of every place, he must himself put it in place B, if he wants it there, for if he were to have put it somewhere else, there is no force capable of removing it from there (*ibid.*, p. 240).

De la Forge's argument is an interesting one. He begins with two premises. The first is the doctrine of divine sustenance, that God must sustain the existence of every body, indeed, of every thing, mind or body, at every moment of its existence. Secondly, de la Forge assumes as a result, it would seem, that God causes motion in the material world by recreating bodies in different places at different times. From this de la Forge draws the conclusion that *only* God can move a body. When God sustains bodies, he must sustain them in *some* place or other; he cannot sustain them everywhere, nowhere, or in any way independently of some place or other. And so causes of motion beside God, causes of motion like our own minds are supposed to be, are neither possible nor needed; if motion and rest are direct results of God's sustenance of the material world, it would seem that there can be no room for other causes.

The position de la Forge is trying to establish here is a variety of occasionalism, and the argument I have sketched is one among many which Descartes's followers used to establish the claim that God is the only genuine cause in the material world, at least.[2] On this view,

* To be presented in an APA symposium on Conceptions of Causality in Early Modern Philosophy, December 29, 1987. Eileen O'Neill will comment; see this JOURNAL, this issue, 580/1.

Many thanks to Roger Ariew for a prompt, careful reading of an earlier draft.

[1] Pierre Clair, ed, *Louis de la Forge: Oeuvres Philosophiques* (Paris: Presses Universitaires de France, 1974). A similar argument is also found in dialogue seven of Nicolas Malebranche, *Dialogues on Metaphysics,* Willis Doney, trans. (New York: Abaris Books, 1980).

[2] For a brief account of occasionalism among 17th Century Cartesians, see ch. 5 of Jean-François Battail, *L'Advocat philosophe Géraud de Cordemoy (1626–1684)* (The Hague: Martinus Nijhoff, 1973). There are a number of varieties of

0022-362X/87/8410/0567$01.40 © 1987 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

causal relations between two bodies, or between a mind and body, are not true causal links, but only occasional causal links which depend for their efficacy on God actually to impart the appropriate motion to the appropriate body. What is especially interesting is that de la Forge starts from what many commentators assume to be genuinely Cartesian doctrines to establish his conclusion. Descartes emphasizes in a number of places that "we have no force through which we conserve ourselves," and so for this we must turn to God, who "continually reproduces us, as it were, that is, conserves us" [Pr I 21].[3] Descartes appeals to this doctrine of divine conservation in proving his laws of nature, both in *Le Monde* and in the *Principia Philosophiae,* arguing that God is the first and continuing cause of motion in the world, and that acting with constancy in preserving his material creation, he must necessarily sustain the world in such a way that certain general constraints on motion are satisfied; quantity of motion is thus conserved, as is motion along a straight path [Pr II 36–42]. The close connection between God's sustenance of the world and his role as cause of motion in the inanimate world have led a number of commentators to see something like de la Forge's view in Descartes, the view that God's role as a cause of motion in the world is inseparable from his role as a sustainer of the world, that God causes motion by creating bodies in different places at different times.[4]

De la Forge's premises seem to belong to Descartes as well. But, if so, then it would appear that, like it or not, Descartes too must be committed to de la Forge's conclusion that God can be the *only* cause of motion in his material world, that, contrary to our "most certain and most evident experience," mind cannot really cause mo-

---

occasionalism. Here I am only concerned with the claim that God is the only genuine cause of motion in the material world.

[3] The numerous references to Descartes's texts will be given in the body, for the most part, using the following abbreviatons. AT: Charles Adam and Paul Tannery, eds., *Oeuvres de Descartes* (Paris: J. Vrin, 1964–1978); CSM: John Cottingham, Robert Stoothoff, and Dugald Murdoch, trans., *The Philosophical Writings of Descartes* (New York: Cambridge, 1985); K: Anthony Kenny, trans., *Descartes: Philosophical Letters* (Minneapolis, Minn: University Press, 1981). The original language text in AT will be cited first, followed by a translation in CSM or K (if available). The one exception to this is the *Principia Philosophiae,* abbreviated 'Pr', and cited by part and section number. The original Latin can be found in AT VIII A, and translations in CSM I.

[4] See e.g., Jean Wahl, *Du rôle de l'idée de l'instant dans la philosophie de Descartes* (Paris: Felix Alcan, 1920); Martial Gueroult, "The Metaphysics and Physics of Force in Descartes," in Stephen Gaukroger, ed., *Descartes: Philosophy, Mathematics and Physics* (Sussex: Harvester Press, 1980), pp. 196–229, esp. 218–220; G. Hatfield, "Force (God) in Descartes' Physics," *Studies in History and Philosophy of Science,* X (1979): 113–140, esp. 127.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tion in the world [AT v 222 (K 235)]. This is the question I would like to examine in this essay. In the end, I shall argue that, when we understand Descartes's doctrine of divine sustenance and of the way God enters the world as a cause of motion, we shall see that, wherever de la Forge's views lead him, Descartes need not be committed to occasionalism, at least not in this way. When we understand just how God causes motion, we shall see that Descartes's God can leave plenty of elbow room for other causes to produce their effects, indeed, produce them as directly as God himself does.

I.

It will be helpful to begin the story with a brief discussion of Descartes's doctrine of divine sustenance. Descartes writes in Meditation III:

> All of the time of my life can be divided into innumerable parts, each of which is entirely independent of the others, so that from the fact that I existed a short time ago, it does not follow that I ought to exist now, unless some cause as it were creates me again in this moment, that is, conserves me [AT VII 49 (CSM II 33)].

Now, Descartes argues, "plainly the same force and action is needed to conserve any thing for the individual moments in which it endures as was needed for creating it anew, had it not existed" [AT VII 49 (CSM II 33)]. Clearly such a power is not in us; if it were, Descartes reasons, I would also have been able to give myself all of the perfections I clearly lack [AT VII 168 (CSM II 118)]. And so he concludes that it must be God that creates and sustains us [AT VII 111, 165, 168, 369–370 (CSM II 80, 116, 118, 254/5); Pr I 21]. This conclusion, of course, holds for bodies as well as for us. It is not just *souls*, but *all* finite things that require some cause for their continued existence. And, as with the idea of ourselves, "when I examine the idea of body, I perceive that it has no power [*vis*] in itself through which it can produce or conserve itself" [AT VII 118 (CSM II 84); cf. AT VII 110 (CSM II 79)]. And so we must conclude that the duration of bodies, too, must be caused by God, who sustains the material world he created in the beginning.

Descartes conceives of God's continual sustenance of his creatures as their *efficient* cause: "I should not hesitate to call the cause that sustains me an efficient cause" [AT VII 109 (CSM II 79)]. But God's causality here is in one respect importantly different from other efficient causes that we are familiar with from our experience. In reply to Gassendi's *Fifth Objections,* Descartes distinguishes between two sorts of efficient causes, a *causa secundum fieri,* a cause of becoming, and a *causa secundum esse,* a cause of being. Roughly speaking, as Descartes undertands the notions, a *causa secundum*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

esse is a cause which must continue to act for its effect to continue, unlike a *causa secundum fieri*, which produces an effect that endures, even after the cause is no longer in operation or even in existence. An architect, thus, is the cause of becoming with respect to a house, as is a father with respect to his son. But Descartes claims

> the sun is the cause of the light proceeding from it, and God is the cause of created things, not only as a cause of becoming, but as a cause of being, and therefore must always flow into the effect in the same way, in order to conserve it [AT VII 369 (CSM II 254/5)].

And, so just as we ordinarily think that the sun must continue its illumination for daylight to persist, so must God continue his activity in order for the world and its motion to be sustained.[5] This continual sustenance is also unlike the more ordinary efficient causes insofar as it requires a kind of power beyond the capacities of created things. Whereas finite things may be able to stand as the efficient causes *secundum fieri* of things in the world, only God, strictly speaking, can stand as their cause *secundum esse*. As we noted earlier, in *Meditation* III Descartes declares that: "plainly the same force and action is needed to conserve any thing for the individual moments in which it endures as was needed for creating it anew, had it not existed" [AT VII 49 (CSM II 33)]. From this Descartes infers that: "it is also one of those things obvious by the light of nature that conservation differs from creation only in reason" [AT VII 49 (CSM II 33)]. That is, the activity and power needed to sustain a thing in its existence is identical to the activity and power necessary to create anything from nothing [cf. also AT VII 165, 166 (CSM II 116, 117)]. Elsewhere he puts the point a bit differently, suggesting that conservation is to be understood as the "continual production of a thing" [AT VII 243 (CSM II 169); cf. Pr II 42], or, more guardedly, suggesting that God *as it were (veluti)* continually reproduces his creatures [Pr I 21; cf. AT VII 110 (CSM II 79)].

In the following section, we shall investigate how Descartes's God causes motion while sustaining the world. But, before turning to that question, I would like briefly to discuss an issue closely related to the questions under discussion here, that of temporal atomism. A number of commentators take Descartes's language quite literally when he says that God must continually recreate his creatures. On their view, Cartesian time must, as a result, be a series of discrete timeless instants, created one after another like the frames of a motion pic-

[5] Descartes does concede, under challenge, that the sun may not be an especially good example of a *causa secundum esse*. See AT III 405, 429 (K 115/6).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ture.[6] Such a view seems inevitably to lead to a position like de la Forge's. The cartoonist creating an animated cartoon can cause his creatures to move only by drawing them in different positions in successive frames; so too for God, it would seem, were we to conceive of him as the grand cartoonist with respect to his creation. In this way, God's sustenance would seem to be inseparable from his role as cause of motion, and all genuine causes of motion other than God would seem to be frozen out.

But it is not at all clear that Descartes held such a position. In a recent study, Jean–Marie Beyssade[7] has argued that Descartes's God sustains the continuously flowing time of our experience. On Beyssade's view, time for Descartes is much like body, infinitely divisible and not composed of any ultimate elements, elements like the durationless temporal atoms are supposed to be. Beyssade does not deny, of course, that Descartes is concerned with timeless instants in a number of important contexts, and, indeed, that he even talks about God conserving bodies as they exist at a given instant [AT XI 44 (CSM I 96); Pr II 39]. But, Beyssade argues, such instants are not, strictly speaking, *parts* of duration. A hunk of extended substance can be divided into innumerable parts. But, for these divisions to be genuine *parts* of a body, they must be extended as well. Points, lines, surfaces, and geometrical objects that lack extension in length, width, and breadth, are not *parts* of a body, but *limits* or *boundaries*. So, Beyssade suggests:

> In the same way, every duration or part of duration contains a before and after . . . ; the instant is its limit or boundary. If we are not mistaken, Descartes always takes this word ['instant'] and its Latin original '*instans*' in the strict sense of a limit (*ibid.*, p. 348; cf. p. 353).

Durations, no matter how small, can be parts of an enduring world, and thus can be candidates for God's sustaining activity. But, although there may be instants *in* duration as boundaries of finite durations, instants, Beyssade suggests, cannot be *parts* of an enduring world; they cannot compose durations, nor can we intelligibly talk about God creating a single instant by itself without creating the duration it serves to bound, any more than we can talk about God creating a two–dimensional surface, a mode of body, without the body that it bounds [AT VII 250/1, 433 (CSM II 174, 292)].

With this in mind, it is easy to see that there is really nothing in

[6] See, e.g., the references cited in note 4.
[7] *La philosophie première de Descartes* (Paris: Flammarion, 1979), pp. 129–143.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
Descartes's texts that unambiguously implies temporal atomism. The idea that all the parts of time are independent, the view we saw earlier in Meditation III, certainly does not; the parts of time in question there might plausibly be read as genuine parts of time, parts with duration, parts which are independent in the sense that God could create any stretch of time without creating preceding or succeeding portions of time. One can give similar readings to other passages in which Descartes talks about the independence of the present time from other moments or moments from one another. Even where Descartes talks of creating the things anew at every moment, even where Descartes makes it clear that God sustains things as they are in a timeless instant, there is no need to attribute temporal atomism to him.[8] To say that God recreates the world at every instant is to say that every instant can be regarded as the beginning, as the boundary of a newly created world. But, although every instant can be regarded as a moment of creation, it does not follow that what is being created is a bare instant or a sequence of bare instants, or that God could create an atemporal instant without creating a duration for that instant to bound.

But, just as Descartes was not committed to temporal atomism, neither was he committed to its denial; I know of few passages that cannot be plausibly interpreted either way. Indeed, I know of no passage to suggest that Descartes was particularly *interested* in the question of temporal atomism, one way or the other. And so it seems improper to argue from Descartes's supposed temporal atomism to the claim that God causes motion through recreating bodies in different positions at different times. If we want to know how God causes motion for Descartes, we should face the question directly.

II.

In presenting his account of God as continual sustainer of the world, Descartes did not think he was telling his readers anything they had not already heard. As far as he was concerned, he was appealing to an old and widely accepted doctrine with which his audience could be expected to be both familiar and generally sympathetic. When in the *Fifth Objections* Gassendi challenged his appeal to a conserving God [AT VII 300 (CSM II 209)], Descartes responded: "When you deny that to be conserved we require the continual influx of a first cause, you deny something that all metaphysicians affirm as obvious" [AT VII 369 (CSM II 254); cf. AT VI 45 (CSM I 133)]. And, in defending himself against Gassendi's criticisms, he seems to have turned di-

[8] For the former formulations see AT VII 49, 109 (CSM II 33, 78/9); for the latter see AT XI 44, 45 (CSM I 96/7), Pr II 39.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

rectly to his copy of St. Thomas Aquinas.[9] God's sustenance of this world of created things is explicitly discussed in the *Summa Theologiae* I, q 104, a 1, and this passage may be the source of Descartes's answer to Gassendi. Like Descartes, Aquinas distinguishes between causes *secundum fieri* and *secundum esse,* and appeals to the same examples Descartes does—the builder of a house, the parent of a child, and the sun as illuminator—to clarify the sense in which God is the cause of the world as enduring [ST I, q 104, a 1 c]. And, although Aquinas does not say exactly that God's activity in sustaining the world is identical with his activity in creating it, many of Descartes's contemporaries would have been happy to agree with Descartes that "this conservation is the very same thing as creation, differing only in reason."[10]

In his monumental commentary on the *Discours,* Étienne Gilson[11] noticed this similarity between Descartes and Aquinas. But Gilson also noted an apparent (and important) difference as well:

> The being of the things Descartes's God conserves is so different from that which St. Thomas's God conserves, that there is a profound difference between their two notions of continual creation. The Thomist God conserves the being of a world of substantial forms and essences. . . . But, on the contrary, in Cartesianism, there are no substantial forms any more . . . (*ibid.*, p. 341).

Gilson goes on to argue that, lacking substantial forms, Descartes, unlike the scholastic, is doomed to a movie–show world of still frames, mocking the continuity of time and motion that the scholastic is genuinely entitled to. Gilson's full argument is too complex to enter into here. But I would like to explore his initial observation a bit.

Descartes does, for the most part, reject substantial forms and this does indeed make a difference, as Gilson emphasizes. But what difference it makes depends upon how the notion is understood, and what it is that takes the place of the absent forms. Now, the notion of a substantial form is a basic notion in Aristotelian thought, and there

[9] In December 1639, Descartes tells Mersenne that he owns "*vne Somme de S. Thomas,*" though it is not altogether clear to me whether this means a copy of Aquinas's *Summa Theologiae* or a *summary of* Aquinas. See AT II 630 (K 69).

[10] Guilelmus Amesius, *Medulla theologica* (1628), quoted in Heinrich Heppe and Ernst Bizer, ed., *Die Dogmatik der evangelisch-reformierden Kirche* (Neukirchen, kreis Moers: Buchhandlung des Erziehungsvereins, 1935), p. 208. For other of Descartes's contemporaries on the question, see Heppe, *Dogmatik,* Locus XII; G. T. Thomson, *Reformed Dogmatics,* Heinrich Heppe, trans. (London: Allen and Unwin, Ltd., 1950), ch. XII; Étienne Gilson, *Index Scholastico-Cartesien* (Paris: J. Vrin, 1979), §§64, 112.

[11] *Discours de la méthode: texte et commentaire* (Paris: J. Vrin, 1967).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

are important differences of conception among scholastic thinkers with regard to that notion. But, to understand the importance of the rejection of substantial forms to Descartes's thought, we must begin with an account of what the notion meant to him.

In very general terms, a substantial form is that which, joined to matter (the *materia prima* of the scholastics, ultimately) results in a complete substance. But, more substantively, substantial form is that from which the characteristic behavior of the various sorts of substances derives. And, so Descartes notes, writing to his then disciple Henricus Regius in January 1642, "they [i.e., forms] were introduced by philosophers to explain the proper action of natural things, of which action this form is the principle and the source" [AT III 506 (K 129)]. Or, as the Conimbrian Fathers wrote in a book Descartes likely learned from as a schoolboy,

> There are individual and particular behaviors [*functiones*] appropriate to each individual natural thing, as reasoning is to human beings, neighing is to horses, heating to fire, and so on. But these behaviors do not arise from matter. . . . Thus they must arise from substantial form (Gilson, *Index*, §209).

More concretely, Descartes views substantial forms as substances of a sort: "by the name 'substantial form' I have understood a certain substance joined to matter, and with it composing something whole that is merely material" [AT III 502 (K 128)]. And the sort of substances they are is mental substance, Descartes thinks, "like little souls joined to their bodies" [AT III 648 (K 135)].[12] And so Descartes characterizes the scholastic account of heaviness which he himself once held as follows:

> But what especially showed that the idea I had of heaviness was derived from that of mind was the fact that I thought that heaviness bore bodies toward the center of the earth as if it contained in itself some knowledge of it [i.e., the center of the earth]. For this could not happen without knowledge, and there cannot be any knowledge except in a mind [AT VII 442 (CSM II 298)].

If substantial forms are supposed to explain the characteristic behavior of bodies of various sorts, then we must be thinking of them as intentional entities, agents of a rudimentary sort, things capable of forming intentions and exercising volition, little souls joined to matter. And so the scholastic doctrine of form and matter is, in a sense,

[12] The remark in question relates to real qualities, strictly speaking, qualities that follow directly from forms. But, in his polemics against the scholastics, Descartes drew no distinction between substantial forms and real qualities.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

just the image of the Cartesian human being, an unextended soul united to extended body and projected out onto the material world. Indeed, Descartes often uses the supposed familiarity of the scholastic model of heaviness (which everyone would have learned at school) to persuade those who have trouble with mind–body interaction on his view that they already understand how interaction is possible; if one can understand the scholastic account of heaviness, then one can understand how the soul can move the body, Descartes reasons, since the two cases are just the same.[13]

It should be evident from Descartes's account of substantial form that he does not reject forms altogether. Given that the human mind is the very model of a form, it is not surprising to find Descartes saying from time to time that the human soul is "the true substantial form of man" [AT III 505]; indeed, it is "the only substantial form" he recognizes [AT III 503; cf. AT IV 346, AT VII 356 (CSM II 246)]. Descartes from time to time also uses scholastic terminology and talks of the soul "informing" the body [AT IV 168 (K 158); AT X 411 (CSM I 40)]. In this sense, Gilson perhaps overestimates the difference between the world Descartes's God sustains and that which Aquinas's sustains, insofar as both contain at least some substantial forms.[14]

But it is significant that Descartes's world has many fewer forms than Aquinas's does, that Descartes rejects all forms but those which pertain to human beings.[15] This raises something of a problem for Descartes, however. The substantial form was that in terms of which the characteristic behavior of a body of a certain sort was to be explained. But, without form, what is to explain why horses neigh and fire heats, why cannon balls fall and smoke rises? In one sense, the replacement for explanation in terms of form is explanation in terms of size, shape, and motion—mechanical explanation. Indeed so, but the story does not end there. In order to explain the behavior of a body (say a cannon ball) mechanistically, we must know more than just the size, shape, and motions of its parts and the surrounding medium; we must also know the relevant laws of motion, how a body as such can be expected to behave, what results when two

[13] For a fuller account of this, see §II of Daniel Garber, "Understanding Interaction: What Descartes Should Have Told Elisabeth," *Southern Journal of Philosophy*, XXI supp. (1983): 15–32.

[14] For a recent discussion of the human soul as substantial form in Descartes, see Marjorie Grene, "Die Einheit des Menschen: Descartes unter den Scholastikern," *Dialectica*, XL (1986): 309–322.

[15] For the still standard account of Descartes' rejection of forms, see Gilson, *Études sur le rôle de la pensée médiévale dans la formation du système cartésien* (Paris: J. Vrin, 1975), pp. 141–190.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

bodies of given sizes, shapes, and motions encounter one another in collision, etc. Descartes replaces the multiplicity of Aristotelian substances, each with its own form and distinct characteristic behavior, with one kind of body which fills the entire universe and behaves everywhere in accordance with the same laws [cf. Pr II 23]. But, in the absence of scholastic substantial forms, Descartes must find some way of explaining the characteristic behavior of material substance, the laws of motion. And it is here that God enters as the "universal and primary [cause of motion], which is the general cause of all motions there are in the [physical] world" [Pr II 36]. God is the cause of motion, what takes the place of the scholastic forms Descartes banished from the inanimate world of nonhuman beings.

But this, of course, leads us back to the question I posed earlier in this paper: How does God cause motion in the world? And how is God's role as cause of motion related to his role as sustainer of body?

To answer this first question, we must, I think, reflect on how souls and the other forms Descartes attributed to the scholastics were thought to cause motion. We must keep in mind here that the issue is under a cloud, so to speak, and it may turn out that, because of an argument like the one de la Forge gave, Descartes is not entitled to hold that the human soul causes motion. But, *prima facie* and despite the doubts of a number of his readers, Descartes certainly thought the question relatively unproblematic. Writing to Arnauld on 29 July 1648, Descartes noted:

> That the mind, which is incorporeal, can impel a body, is not shown to us by any reasoning or comparison with other things, but is shown daily by the most certain and most evident experience. For this one thing is among the things known *per se*, which we obscure when we try to explain through other things [AT V 222 (K 235)].

The mind can cause motion in a body, Descartes holds; it is something we know through experience directly, something we cannot explain in other terms. Insofar as we comprehend it, it is because "we have within us certain primitive notions, on the model of which we form all our other knowledge," Descartes explains to Elisabeth [AT III 665 (K 138)]. And, Descartes goes on to explain, we understand the schoolman's substantial forms to work in exactly the same way; indeed, as I noted earlier, the understanding we have of forms derives from the notions we were given to understand how human mind works on body.

Descartes is not very informative about just how mind (or form) moves body; on his view, there is not much that *can* be said, other than to direct our attention to the experience we all have that is

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

supposed to make it all clear.[16] But, although there is not much we can say, there is no confusing the sense in which mind causes motion in a body with the way God sustains the body that mind supposedly moves. One of the axioms Descartes uses in the geometrical presentation of his arguments appended to the *Second Replies* reads as follows: "It is greater to create or conserve a substance, than it is to create or conserve the attributes or properties of a substance" [AT VII 166 (CSM II 117)]. The passage is not without its difficulties.[17] But the clear sense is that Descartes wants to distinguish causes that change the modes or properties of a thing (modal causes, as I shall call them) from causes that create or sustain the very being of a substance (substantial causes, perhaps). God, sustaining the world, is clearly a substantial cause. But minds are clearly not; insofar as they cause changes in the motion of bodies, they at best can count as modal causes. And, insofar as substantial forms are understood on the model of souls acting on bodies, Descartes would have had little trouble classifying them, with minds, as modal causes; they are causes of motion, a mode in bodies assumed to be sustained by the divine sustainer who is the unique substantial cause.

God enters Descartes's physics to do the business substantial forms did in the Aristotelian system, as he understood it, to cause bodies to behave in their characteristic ways. And, I claim, when doing the business of forms, Descartes's God is understood to cause motion in just the way forms were taken to do it, that is, on Descartes's account, in just the way that we do it: by way of an impulse that moves matter in a way that we can comprehend only through immediate experience. This is not at all clear as late as 1644 when, in proving his laws of nature in the *Principia,* Descartes's account of God as cause of motion is deeply (and obscurely) intertwined with his account of God as sustainer of the world [Pr II 36–42]. But, by April 1649, Descartes wrote to Henry More:

> Although I believe that no mode of acting belongs univocally to God and to his creatures, I confess nevertheless, that I can find no idea in my mind which represents the way in which God or an angel can move matter, which is different from the idea that shows me the way in which I am conscious that I can move my own body through my thought [AT V 347 (K 252)].

[16] This, in any case, is what he insisted on in writing to Elisabeth. See AT III 663–668, 690–695 (K 137–143). What exactly he meant here is not entirely clear.

[17] The passage raises an obvious question about the relation between an attribute and a substance, a question the young Burman raised to Descartes in coversation. See AT V 154, and John Cottingham, trans. and ed., *Descartes' Conversation with Burman* (New York: Oxford, 1976), pp. 15, 77–80.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

And, so Descartes suggests to More, God is conceived to move bodies in just the way we do, using the same primitive notion we use to understand how we move our own bodies.

If this is how we conceive of God as a cause of motion, then, it would seem, we are conceiving of him as a *modal cause* when it comes to motion. Conceived as such, there would appear to be a distinction between God as sustainer of the world, a substantial cause keeping things in existence, and God as cause of motion, a modal cause causing bodies to have the particular motion they have, determining, at least in part, their modes. The difference between these two roles God plays for Descartes comes out again in the correspondence with More. In his letter of 5 March 1649, More asked Descartes if "matter, whether we imagine it to be eternal or created yesterday, left to itself, and receiving no impulse from anything else, would move or be at rest?" [AT v 316]. Descartes answers in his letter of August 1649: "I consider 'matter, left to itself, and receiving no impulse from anything else' as plainly being at rest. But it is impelled by God, conserving the same amount of transference in it as he put there from the first" [AT v 404 (K 258)]. The picture that comes through here is a simple one. Bodies can be conserved with or without the divine impulse. Without the impulse, they are at rest; with it, in motion. God's conservation of body seems separable from his role as cause of motion, and as cause of motion, he seems to act as we would in the circumstances; God's motion seems to result from a divine act of will, a divine shove.

III.

Now that we understand something of how God causes motion, we can return to the question originally posed and offer an answer.

As de la Forge construed Descartes's views on God, continual recreation, and God's role as cause of motion, Descartes seems pushed inevitably toward occasionalism and the view that God is the only genuine cause of motion in the world; if God causes motion by recreating bodies in different places at different times, then there seems to be no room for finite causes to act. But by now it should be clear why Descartes need not be committed to such a view.

I have argued that, for Descartes, God enters as a cause of motion in order to replace the scholastic's substantial forms, and, in that role, he can (and, in the More letters, at least, *is*) construed as acting in just the way forms were thought to cause motion, that is, in just the way *we* cause motion. As such, God both sustains bodies in their being and sustains bodies in their motion. But, it is important to note, *these two activities seem to be quite distinct;* in the one case, God is acting as a modal cause, in the other, as a substantial cause.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

This is an extremely important observation. There is no substantial cause but God, nor can there be, since no other being has the ability to create and sustain the universe. But, although God is a modal cause with respect to motion, there is no reason to hold that God is the *only* such cause. God is conceived to act as we do in causing motion; just as the finite cause of motion does not exclude others, so the fact that God causes motion does not seem to exclude other causes. This seems true even when we are talking about causing motion *in the same body*. Just as two human beings can exert their contrary impulses on the same bit of matter, so can we impose an impulse contrary to the one God imposes. Indeed, we do so every time we lift a stone, on which God is imposing an impulse to move toward the center of the earth.

And so Descartes would have to agree with de la Forge that God cannot sustain bodies that are in no place at all or in indeterminate places; the very possibility is absurd. But, I think Descartes might insist, although God sustains bodies that *have* place, it is not the act of sustaining them that *gives* them place. What gives them place and the motion that puts them in different places at different times is impulse or the lack thereof, a cause quite distinct from that by which bodies are sustained. These impulses may come from God himself, but they might come from other causes, like our own minds [cf. AT v 403/4 (K 257)]. And, when they come from God, they are not to be identified with the cause by which he sustains the bodies he moves.

There are a number of important questions relevant to the topic at hand which space will not permit us to discuss. Most importantly, it would be valuable to discuss the relations between the conception of motion and its divine cause which I have been developing with the discussion of motion and rest and their laws in part II of the *Principles* and in ch. 7 of *The World*—the sense in which motion and rest are distinct and the sense in which they are not, the sense in which motion and rest are states, and the way in which motion and rest give rise to forces that come into play at the time of collision. My story will not be complete until we see how the way in which Descartes's immutable God causes motion leads him to the conception of motion (and its associated forces and laws) which underlies his program in natural philosophy.

But, incomplete as my preliminary sketch of Descartes's position may be, it allows us to see one important feature that differentiates Descartes's metaphysic of motion and his use of God as cause of motion from that of his avowedly occasionalist followers. What lies behind occasionalism as advanced by de la Forge and by many Cartesians of his generation is a deep worry about causality in the world

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of finite things; what comes up again and again is the view that finite things are incapable of any genuine causal efficacy, that producing an effect is beyond the power of any finite thing. God enters as the *only* being capable of producing any change in the world.[18] Descartes's view is quite different. Descartes *never* rejects finite causes as such; indeed, it is on the model of one particular finite cause, us, that *all* causes are understood, conservation excepted.[19] When God enters as a cause of motion, it is simply on account of the fact that some finite causes needed to do the job are not available. But, even when God undertakes this task, it seems to me that Descartes can quite well hold that finite causes of motion are in no way squeezed out. Mind, indeed, can remain as direct a cause of motion for Descartes as God himself.

DANIEL GARBER

University of Chicago

DIVINE IMPULSE, VOLUNTARY MOTION, AND INTELLIGIBLE CAUSAL MODELS*

In his earlier paper, "Mind, Body, and the Laws of Nature in Descartes and Leibniz,"[1] Daniel Garber cleared Descartes of Leibniz's charge that mind–body interaction violated the conservation law. This defense rested on showing the specific way Descartes derived the laws of nature from God. In particular, it was argued that the laws of motion govern the way God acts upon the material world, viz. through continuous creation. Near the end of the paper, Garber raised Malebranche's challenge: If Descartes is committed to the continuous creation doctrine, then he ought to hold the view that, at each instant, God is the only source of any change of place of bodies. Thus, minds, for example, cannot cause motion.

In "How God Causes Motion: Descartes, Divine Sustenance, and Occasionalism," Garber takes up this concern, focusing on Louis de la Forge's version of the argument. Garber reasons that Descartes's

[18] See especially Nicolas Malebranche, *De la récherche de la vérité,* bk. VI pt. II, ch. III, and the *XVe Éclaircissement.*

[19] See Garber, "Understanding Interaction," *op. cit.*

* Abstract of a paper to be presented in an APA symposium on Conceptions of Causality in Early Modern Philosophy, December 29, 1987, commenting on a paper by Daniel Garber, this JOURNAL, this issue, 567–580.

[1] *Midwest Studies in Philosophy,* VIII (1983): 105–133.

0022-362X/87/8410/0580$00.50 © 1987 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

views imply occasionalism, only if the way God sustains the world from instant to instant is the means by which he causes motion in the inanimate world.

The main argument against de la Forge is right on target: the Cartesian texts *do* imply a distinction between the "impulse" by means of which God moves bodies, and the activity by which God sustains bodies in their being. Although God alone can sustain the world, divine impulse is understood to work just like a finite cause of motion, the impulse of the will. And, since finite causes of motion do not exclude others, divine impulse, then, leaves room for other causes of motion contrary to de la Forge, et al.

Although this defense of Descartes against his occasionalist critics is a position Descartes himself would have, or at least should have, held, the following question still remains: Precisely *how* does God (or the will) cause motion?

In my paper, I shall examine the concept of "impulse," which Descartes uses to characterize the ways God and the mind cause motion in bodies. I shall sketch the connections between "impulse," "impressed force," and "impetus," as they appear in medieval views about projectile motion. The relation of these concepts to that of the "influence or inflowing of species" will also be traced. It will become clear that, by the beginning of the seventeenth century, 'impulses' and 'inflowings of similitude, impression, or species' denote communications of properties or powers such that: (1) there is no contact of substances, (2) the cause loses nothing of its virtue (or power), and (3) the effect exists only so long as there is continual activity of the cause. Examples of such impulses or influences include the communication of alchemical forces like magnetic power, and astrological forces like heaviness. Descartes rejects the account of these phenomena in terms of impulses or inflowings, in favor of purely mechanistic explanations. But he retains the impulse picture of how God and the mind act on matter.

I shall argue, however, that, for Descartes, the impulse and influence picture do not constitute an intelligible causal model. Rather, mechanism alone provides a causal paradigm that is cognitively accessible to us. In the end, divine impulse and voluntary motion remain occult.

In conclusion, I shall raise some questions about motive force and the degree of causal efficacy which Descartes attributes to his "secondary causes."

EILEEN O'NEILL

Queens College and The Graduate Center
City University of New York

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

The Department of Philosophy of Rice University is pleased to announce the following appointments: C. Kenneth Waters as Assistant Professor, and Karey Harrison as Visiting Assistant Professor. Paul Warren has been reappointed as Visiting Assistant Professor.

The Department of Philosophy of the University of Vermont is pleased to announce the appointment of David Christensen (UCLA) as Assistant Professor, effective September 1, 1987.

The Department of Philosophy of California State University/Northridge is pleased to announce the appointment of Takashi Yagisawa as Assistant Professor.

The Department of Philosophy of Ohio State University is pleased to announce that Simon Blackburn (Oxford) will be visiting Distinguished Professor during the Spring quarter, 1988.

The Department of Philosophy of the University of Pennsylvania is pleased to announce the following appointments, effective July 1, 1987: Zoltan Domotor and Scott Weinstein as Professors of Philosophy, and Gary Hatfield as Associate Professor of Philosophy. The Department is also pleased to announce the appointment of Alexander Nehamas as Professor of Philosophy as of July 1, 1986.

The Department of Philosophy of Brown University is pleased to announce that Jaegwon Kim will join the department as the William Herbert Perry Faunce Professor of Philosophy, effective July 1, 1987.

The Society for Women in Philosophy is pleased to announce that its 1987 Honorary Woman of the Year is Elizabeth Flower (Pennsylvania). Elizabeth Beardsley (Temple) was its first honoree in 1984; she was followed by Marjorie Grene (California/Davis) in 1985 and Mary Mothersill (Barnard) in 1986.

The APA Committee on the Status of Women is pleased to announce publication of the *Feminism and Philosophy Newsletter*. The newsletter will provide information about recent work in feminist philosophy, list new publications, and publish book reviews. It will include discussions of how to integrate this material into traditional philosophy courses. The editor, Nancy Tuana (Texas/Dallas), invites contributions: literature overviews, book reviews, suggestions for curriculum revisions, and discussions of feminist pedagogical methods are especially welcome. Manuscripts, proposals, suggestions, and inquiries may be addressed to Nancy Tuana, Editor, Arts and Humanities, JO 31, University of Texas/Dallas, Richardson, TX 75083-0688.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The editors are pleased to announce a new book series, *Soundings,* to be published by Notre Dame University Press. It will offer scholarly, graduate-level analyses of social and moral issues in business or economics. The scope of the series is interdisciplinary, encompassing work in the social sciences, philosophy, economics, theology, business, political science, and public policy. Thomas Donaldson (Loyola) will edit the series; manuscripts and proposals should be addressed to him at the Department of Philosophy, Loyola University of Chicago, 6525 N. Sheridan Road, Chicago, IL 60626.

The Center for Philosophy of Religion at the University of Notre Dame announces one senior and two junior fellowships for the 1988–89 academic year. The fellowships will enable holders to pursue research and writing projects in Christian philosophy and the philosophy of religion. Each fellow is asked to teach one class a semester. Those with sabbatical support are welcome to apply. Applicants should submit a complete *curriculum vitae,* three letters of recommendation, a statement (double-spaced) describing the prospective project, and one published or unpublished paper. Submissions and inquiries may be addressed to Alvin Plantinga, Director, Center for Philosophy of Religion, 330 Decio Hall, P.O. Box 1068, Notre Dame, IN 46556.

The North American Division of the Schopenhauer Society announces a call for papers for its Jubilaemskongress honoring the 200th Anniversary of Schopenhauer's birth. The congress will be held in conjunction with the Eastern Division meeting of the APA in December 1988. Papers on any aspect of Schopenhauer's thought are welcome; deadline for submissions is April 15, 1988. Submissions and inquiries should be addressed to David Cartwright, Department of Philosophy and Religious Studies, University of Wisconsin, Whitewater, WI 53190.

The editors welcome the announcement of a new quarterly journal, *Cultural Dynamics,* to be published by Brill (Leiden & New York). *Cultural Dynamics* will be a forum both for theoretical and empirical studies focusing on processes, changes, evolutions, and histories of knowledge and of social and cultural phenomena. The editors of *Cultural Dynamics* welcome submissions, which should be addressed to Rik Pinxten, CD, Blandijnberg 2, 9000 Gent, Belgium. Subscription information may be obtained from Brill Publishers, P.O. Box 9000, 2300 PA Leiden, Netherlands.

The Department of Philosophy of the University of Hawaii/Manoa and the Society for Asian and Comparative Philosophy are pleased to announce that under their sponsorship, and in cooperation with the East–West Center, a Sixth East–West Philosopher's Conference will be held in Honolulu July 30–August 12, 1989. The general theme of the conference will be "Culture and Modernity," focusing on the principles of cultural change and the meaning of "modernity" and "post-modernity" in both Western and Asian contexts within religion and art, law and ethics, science and technology. For information, write: Director, Sixth East–West Philosopher's Conference, Department of Philosophy, 2530 Dole Street, Honolulu, HI 96822.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

A conference on the topic of Christian/theistic philosophy will be held at the University of Notre Dame from February 25–27, 1988. Principal speakers will include Alan Donagan (California Institute of Technology), Jorge Garcia (Notre Dame), Norman Kretzmann (Cornell), Christopher Menzel (Texas A & M), Richard Purtrill (Western Washington), Eleonore Stump (Virginia Polytechnic Institute), Merold Westphal (Hope College), and Nicholas Wolterstorff (Calvin College). The conference is sponsored by the Notre Dame Center for Philosophy of Religion. For further information, write to Alvin Plantinga, Center for Philosophy of Religion, 330 Decio Hall, P.O. Box 1068, University of Notre Dame, Notre Dame, IN 46556.

Siena College announces a call for papers for its third annual multidisciplinary conference, this year on the fiftieth anniversary of World War II. While papers on Czechoslovakia and the Munich Conference will be particularly appropriate, papers dealing with Fascism and Nazism, the war in Asia, Spain, literature, art, film, diplomacy, political and military history, popular culture, and women's and Jewish studies are also welcomed. Deadline for submissions is December 15, 1987. Inquiries and submissions may be addressed to Thomas O. Kelly II, Department of History, Siena College, Loudonville, NY 12211.

A conference celebrating the 350th anniversary of Descartes's *Discourse on Method* and the attached *Essays* will be held in San Jose, California, on April 15–17, 1988, featuring both invited and contributed papers. A contributed paper should take between 25 and 40 minutes to read. It should focus on some aspect of the publications of 1637 or of the famous "tree of philosophy," consisting of Descartes's metaphysics, physics, mechanics, medicine, and morals. We expect the conference to clarify Descartes's method and its applications, and, also, with the benefit of 350 years of hindsight, to contribute to the assessment of his philosophy. The planning committee must receive contributed papers by December 15. Please send papers and other communications to the conference organizer: Stephen Voss, Department of Philosophy, San Jose State University, San Jose, CA 95192.

Harvard University's Program in Ethics and the Professions invites applications for Fellowships in Professional Ethics. At least four fellowships will be awarded in 1988–89 to outstanding teachers and scholars who wish to develop their competence to address ethical issues in schools of business, government, law, and medicine. Fellows will participate in the seminar of the program, attend courses in one of the professional schools or in the Graduate School of Arts and Sciences, and conduct their own research on professional ethics. The duration of the fellowship is eleven months, September–July. Applicants should hold a postgraduate degree in business, government/public policy, law, or medicine, or a doctorate in philosophy, political theory, or theology. Application deadline is December 31, 1987. For further information, call or write Dr. Valerie Abrahamsen, Administrator, The Program in Ethics and the Professions, Harvard University, 79 J.F.K. Street, Cambridge, MA 02138.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

AMERICAN PHILOSOPHICAL ASSOCIATION
EASTERN DIVISION
EIGHTY-FOURTH ANNUAL MEETING

SHERATON-CENTRE HOTEL, NEW YORK CITY
DECEMBER 27–30, 1987

PROGRAM

SUNDAY, DECEMBER 27

1:00 P.M.–10:00 P.M. REGISTRATION, *Sheraton Exhibit Center Foyer*
1:00 P.M. EXECUTIVE COMMITTEE MEETING, *St. James B*
8:00 P.M. SOCIETY OF CHRISTIAN PHILOSOPHERS, *Royal Ballroom B*
8:00 P.M. CONFERENCE OF PHILOSOPHICAL SOCIETIES *Versailles Ballroom*
8:00 P.M. SOCIETY FOR CLASSICAL REALISM, *Senate*
8:00 P.M. THE SOCIETY OF PHILOSOPHERS AT WORK IN THE WORLD *Regency Foyer*

MONDAY, DECEMBER 28

9:00 A.M.–7:00 P.M. REGISTRATION, *Rear Main Lobby*
9:00 A.M.–5:00 P.M. BOOK DISPLAYS, *Sheraton Exhibit Center*

CHILD CARE arrangements may be made on an individual basis with a commercial firm. See the September or November issue of the APA *Proceedings* for further details.

PLACEMENT INFORMATION

For interviewers, 8:00 A.M.–7:00 P.M., *Versailles Terrace*
For candidates, 9:00 A.M.–7:00 P.M., *Princess Ballroom*

MORNING (9:00–11:00)

Group Meetings (see Appendix for Programs)

9:00 NORTH AMERICAN NIETZSCHE SOCIETY, *Royal Ballroom B*
9:00 RADICAL PHILOSOPHY ASSOCIATION, *Versailles Ballroom*
9:00 SOCIETY FOR WOMEN IN PHILOSOPHY, *Georgian Ballroom A*
9:00 ASSOCIATION FOR SYMBOLIC LOGIC, *Georgian Ballroom B*
9:00 INTERNATIONAL PHILOSOPHERS FOR THE PREVENTION OF NUCLEAR OMNICIDE and CONCERNED PHILOSOPHERS FOR PEACE, *Regency Ballroom*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Monday, December 28, Morning

9:00 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Regency Foyer*
9:00 INTERNATIONAL THOMAS AQUINAS SOCIETY, *Commonwealth*
9:00 SOCIETY FOR THE STUDY OF PROCESS PHILOSOPHIES, *Senate*
9:00 SOCIETY FOR PHILOSOPHY AND TECHNOLOGY, *Dominion A*
9:00 NORTH AMERICAN SOCIETY FOR SOCIAL PHILOSOPHY *Dominion B*
9:00 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Province*
9:00 BERTRAND RUSSELL SOCIETY, *Embassy*
9:00 APA COMMITTEE ON INTERNATIONAL COOPERATION, SOCIEDAD FILOSOFICA IBEROAMERICANA, and SOCIETY FOR IBERIAN AND LATIN AMERICAN THOUGHT, *Consulate*
9:00 COUNCIL FOR PHILOSOPHICAL STUDIES, *St. James A*

LATE MORNING (11:15–1:15)

11:15 SOCIETY FOR THE ADVANCEMENT OF AMERICAN PHILOSOPHY *Versailles Ballroom*
11:15 AMERICAN SOCIETY FOR VALUE INQUIRY, *Regency Ballroom*
11:15 SARTRE CIRCLE, *Regency Foyer*
11:15 INTERNATIONAL SOCIETY FOR NEOPLATONIC STUDIES, *Diplomat*
11:15 ASSOCIATION FOR PHILOSOPHY AND LIBERATION *Commonwealth*
11:15 JASPERS SOCIETY OF NORTH AMERICA, *Senate*
11:15 CONCERNED PHILOSOPHERS FOR PEACE, *Dominion A*
11:15 MOORE SOCIETY, *Embassy*
11:15 WORLD PHENOMENOLOGY INSTITUTE, *Monarch*

AFTERNOON

2:00–5:00 Session I

A. Symposium: Epistemology and Philosophy of Mind *Royal Ballroom A*
   Chair: Robert McCauley
   Speakers: Alvin I. Goldman and Patricia S. Churchland
   Commentator: George Bealer

B. Symposium: Husserl and Frege, *Imperial Ballroom B*
   Chair: Jitendra Mohanty
   Speakers: Robert Sokolowski and Ronald McIntyre
   Commentator: Donn Welton

C. Invited Papers: *Versailles Ballroom*
   1. Philosophy and Physics
      Chair: Peter Machamer
      Sylvan S. Schweber, Effective Field Theories and Hierarchical Descriptions
      Commentator: James Cushing

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

2. Skepticism, Old and New
   Chair: John Heil
   Julia Annas, Skepticism Old and New
   Commentator: Richard H. Popkin

D. Colloquium: Logic, Probability, and Methodology
Royal Ballroom B

1. *Chair:* Henry E. Kyburg, Jr.
   John M. Vickers, Logic and Thorough Probabilism
   Commentator: Deborah Mayo
2. *Chair:* Rod Stewart
   Bernard Kobes, Prototype Theory without Psychologism
   Commentator: Richard Grandy
3. *Chair:* William Harper
   W. J. Talbott, Reflections on Two Principles of Bayesian Epistemology
   Commentator: Teddy Seidenfeld

E. Colloquium: Time, *Georgian Ballroom A*

1. *Chair:* Charles Sherover
   Mary Jeanne Larrabee, Time and Spatial Models: Temporality in Husserl
   Commentator: Peter McCormick
2. *Chair:* Donald Baxter
   Michelle Beer, Temporal Indexicals and the Passage of Time
   Commentator: Jane Duran
3. *Chair:* Patrick Heelan
   James Plecha, Is Modern Physics Committed to an Ontology of Time?
   Commentator: Kristin Shrader-Frechette

F. Colloquium: Experience, Science and Religion. *Regency Foyer*

1. *Chair:* Jaimie Ferreira
   Gary Legenhausen, Religious Experiences Do Not Provide Prima Facie Justification for Religious Beliefs
   Commentator: Louis Pojman
2. *Chair:* James Harris
   Paul Draper, An Evidential Problem of Evil
   Commentator: Billy Joe Lucas
3. *Chair:* Ernan McMullin
   William L. Craig, The Anthropic Principle versus Divine Design
   Commentator: John Earman

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Monday, December 28, Afternoon, Session I

G. Special Invited Panel: On the U.S. Constitution (Bicentennial) *Regency Ballroom*

Chair: Martin Golding
Panelists: Bruce A. Ackerman, Andrew Reck, and Frank I. Michelman

H. Special Symposium Arranged by the APA Committee on International Cooperation, *Commonwealth*
Topic: Philosophy in Austria
I. *Chair:* Roderick Chisholm
Rudolf Haller, Austrian Philosophy
Reinhard Kamitz, Formal Logic and Intuitive Inference
II. *Chair:* Keith Lehrer
Edgar Morscher, Ontology in Elementary Logic
Paul Weingartner, Philosophy of Science

I. Special Symposium Arranged by the APA Committee on the Status of Women, *Monarch*
I. Business Meeting (Alison M. Jaggar presiding)
II. *Topic:* Feminism and Environment
Chair: Alison M. Jaggar
Speakers: Irene Diamond, Ynestra King, and Karen J. Warren
Commentator: James Cheney

Group Meetings (See Appendix for Programs)

2:00 ASSOCIATION FOR SYMBOLIC LOGIC, *Georgian Ballroom B*
2:00 AMERICAN ASSOCIATION OF PHILOSOPHY TEACHERS, *Diplomat*
2:00 AMERICAN ASSOCIATION FOR THE PHILOSOPHIC STUDY OF SOCIETY and INSTITUTE FOR HUMANE STUDIES, *Senate*
2:00 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING *Consulate*

LATE AFTERNOON (5:15–7:15)

5:15 RADICAL PHILOSOPHY ASSOCIATION, *Versailles Ballroom*
5:15 SOCIETY FOR WOMEN IN PHILOSOPHY, *Regency Ballroom*
5:15 PEIRCE SOCIETY, *Regency Foyer*
5:15 SOCIETY FOR THE PHILOSOPHICAL STUDY OF THE PARANORMAL *Diplomat*
5:15 INTERNATIONAL SOCIETY FOR CHINESE PHILOSOPHY *Commonwealth*
5:15 GANDHI-KING SOCIETY, *Senate*
5:15 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY *Consulate*
5:15 SOCIETY FOR THE STUDY OF ETHICS AND ANIMALS, *Dominion A*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

5:15 SOCIETY FOR CLASSICAL REALISM, *Embassy*
5:15 SOCIETY FOR THE PHILOSOPHICAL STUDY OF MARXISM, *Monarch*
5:15 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Province*
5:15 SOCIETY FOR BUSINESS ETHICS, *Dominion B*
5:15 SOCIETY OF PHILOSOPHY JOURNAL EDITORS, *Royal Ballroom B*

EVENING (7:30–10:30)

7:30 NORTH AMERICAN KANT SOCIETY, *Versailles Ballroom*
7:30 AMERICAN SOCIETY FOR VALUE INQUIRY, *Regency Ballroom*
7:30 SANTAYANA SOCIETY, *Regency Foyer*
7:30 LEIBNIZ SOCIETY OF AMERICA, *Senate*
7:30 SOCIETY FOR IBERIAN AND LATIN AMERICAN THOUGHT, *Consulate*
7:30 ASSOCIATION FOR SYMBOLIC LOGIC COUNCIL MEETING *Dominion A*
7:30 ASSOCIATION FOR PHILOSOPHY AND PUBLIC AFFAIRS *Dominion B*
7:30 PERSONALIST DISCUSSION GROUP, *Embassy*
7:30 PHILOSOPHERS FOR SOCIAL RESPONSIBILITY, *Monarch*
7:30 WORLD PHENOMENOLOGY INSTITUTE, *Province*
7:30 SOCIETY FOR ANCIENT GREEK PHILOSOPHY, *Royal Ballroom B*

8:00 RECEPTION, *Imperial Ballroom*

TUESDAY, DECEMBER 29

9:00 A.M.–5:00 P.M. REGISTRATION, *Rear Main Lobby*
9:00 A.M.–5:00 P.M. BOOK DISPLAYS, *Sheraton Exhibit Center*

PLACEMENT INFORMATION

For interviewers, 8:00 A.M.–5:00 P.M., *Versailles Terrace*
For candidates, 9:00 A.M.–5:00 P.M., *Princess Ballroom*

MORNING

9:00–11:00 Session II

A. Symposium: Conceptions of Causality in Early Modern Philosophy, *Royal Ballroom A*
Chair: Lynn Joy
Speaker: Daniel Garber
Commentator: Eileen O'Neill

B. Symposium: Liberties, Wrongs, and Representation *Royal Ballroom B*
Chair: Joseph Flay
Speaker: Jay Drydyk
Commentator: Martin de Nys

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Tuesday, December 29, Morning, Session II

C. Invited Paper: Ethics, *Versailles Ballroom*
Chair: Alan Goldman
John McDowell, Might There Be External Reasons?
Commentator: Thomas Scanlon

D. Colloquium: American Philosophy, *Imperial Ballroom B*
1. *Chair:* Beth Singer
Victorino Tejera, Kulick's Misrepresentations of American Philosophy
Commentator: Roland Garrett
2. *Chair:* James Gouinlock
Mark Mendell, Dewey and the Fiction of Corporate Legal Personality
Commentator: Raymond Boisvert

E. Colloquium: Ethics, *Georgian Ballroom A*
1. *Chair:* Terrance McConnell
Earl Conee, Why Moral Dilemmas Are Impossible
Commentator: Peter Vallentyne
2. *Chair:* Gregory Trianowski
Ferdinand Schoeman, Cognitive Limits and Moral Heuristics: Against Autonomy in Ethics
Commentator: Allen Fuchs

F. Colloquium: Friendship, *Regency Foyer*
1. *Chair:* Edward Sankowski
David B. Annis, Emotion, Love and Friendship
Commentator: Sander H. Lee
2. *Chair:* Jennifer Whiting
David O'Connor, Two Ideas of Friendship
Commentator: Ann Wiles

G. Colloquium: Hegel and Sartre, Aesthetics, *Regency Ballroom*
1. *Chair:* William McBride
Robert R. Williams, Intersubjectivity and First Philosophy: Sartre's Appropriation of Hegel
Commentator: Stephen Watson
2. *Chair:* Richard E. Hart
George Bailey, Forging Metaphor
Commentator: Gary Shapiro

H. Special Panel Arranged by the APA Committee on the Teaching of Philosophy, *Monarch*
Topic: How to Improve Your Teaching
Chair: Karen Hanson
Panel: Steven M. Cahn, Chad D. Hansen, and Michael Hooker

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Group Meetings (See Appendix for Programs)

9:00 ASSOCIATION FOR SYMBOLIC LOGIC, *Georgian Ballroom B*
9:00 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Diplomat*
9:00 SOCIETY FOR THE PHILOSOPHY OF SEX AND LOVE, *Commonwealth*
9:00 SOCIETY FOR THE PHILOSOPHY OF CREATIVITY, *Senate*
9:00 SARTRE CIRCLE, *Embassy*

11:15 RADICAL PHILOSOPHY ASSOCIATION, *Versailles Ballroom*

11:30 A.M.–1:30 P.M. BUSINESS MEETING, *Imperial Ballroom A*

AFTERNOON

2:15–4:15 Session III

A. Symposium: Relevant Alternatives and Demon Skepticism
Royal Ballroom A

Chair: Paul Moser
Speaker: Bredo C. Johnsen
Commentator: Jonathan Adler

B. Symposium: Public Policy and the Allocation of Scarce Medical Resources, *Royal Ballroom B*
Chair: Dan Brock
Speaker: Richard Barber
Commentator: Norman Daniels

C. Invited Paper: Vico on History and Knowledge
Imperial Ballroom B

Chair: Richard J. Bernstein
Donald P. Verene, Imaginative Universals and Narrative Truth
Commentator: Alasdair MacIntyre

D. Colloquium: Epistemology, *Versailles Ballroom*
1. *Chair:* C. Thomas Powell
Steven F. Savitt and Morrie Lipson, Information, Reliability, and Knowledge
Commentator: Jonathan Vogel
2. *Chair:* Michael Losonsky
William E. Morris, Information and the Ebb of Knowledge: Fred Dretske's Information-theoretic Epistemology
Commentator: Michael Hand

E. Colloquium: Whitehead and Process Philosophy
Georgian Ballroom A

1. *Chair:* George Allan
Lewis S. Ford, Hartshorne on Perishing in Whitehead
Commentator: Bowman Clarke

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Tuesday, December 29, Afternoon, Session III–E

2. *Chair:* William Garland
   George R. Lucas, Analytic Themes in Whitehead's Metaphysics
   Commentator: Donald Sherburne

F. Colloquium: Plato's *Phaedrus, Regency Foyer*

1. *Chair:* Cynthia Hampton
   Catherine M. Culver, Method and Madness in Plato's *Phaedrus*
   Commentator: P. Christopher Smith

2. *Chair* Anne Ashbaugh
   David A. White, Division and Nature in Plato's *Phaedrus*
   Commentator: Mark McPherran

G. Colloquium: History of Philosophy, *Regency Ballroom*

1. *Chair:* Arthur S. McGrade
   Clyde Lee Miller, Perception and Coniectura in Nicholas of Cusa
   Commentator: Fred Purnell

2. *Chair:* Harry van der Linden
   Rosamond Rhodes, Obligation and Assent in Hobbes's Moral Philosophy
   Commentator: Tommy Lott

H. Special Symposium Arranged by the APA Committee on Blacks in Philosophy, *Diplomat*
   Topic: Civil Rights Legislation and Black Programs
   Chair: Thomas Slaughter
   Bernard Boxill, Has Civil Rights Legislation Reached the Point of Diminishing Returns?
   Commentator: Blanche Curry

I. Special Symposium Arranged by the APA Committee on Teaching Philosophy at Two-year Colleges, *Province*
   Topic: The Importance of Philosophy for Two-year Programs
   Chair: Neil Rossman
   Speaker: Eric Lindermayer
   Commentator: Tziporah Kasachkoff

Group Meetings (See Appendix for Programs)

2:15 ASSOCIATION FOR SYMBOLIC LOGIC, *Georgian Ballroom B*
2:15 SOREN KIERKEGAARD SOCIETY, *Commonwealth*
2:15 CHAIRS' CAUCUS, *Senate*
2:15 AMERICAN SOCIETY FOR VALUE INQUIRY, *Monarch*

4:30 Presidential Address: *Imperial Ballroom A*
   Introduction by Richard J. Bernstein
   Jonathan Bennett, Thoughtful Brutes

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

EVENING (6:30–9:30 P.M.)

Group Meetings (See Appendix for Programs)

6:30 RADICAL PHILOSOPHY ASSOCIATION and SOCIETY FOR THE PHILOSOPHICAL STUDY OF MARXISM, *Versailles Ballroom*
6:30 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Regency Ballroom*
6:30 CONFERENCE OF PHILOSOPHICAL STUDIES, *Regency Foyer*
6:30 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY *Commonwealth*
6:30 AMERICAN ASSOCIATION FOR THE PHILOSOPHIC STUDY OF SOCIETY and INSTITUTE FOR HUMANE STUDIES, *Senate*
6:30 JASPERS SOCIETY OF NORTH AMERICA, *Consulate*
6:30 ASSOCIATION FOR SYMBOLIC LOGIC COUNCIL MEETING *Dominion A*
6:30 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING *Embassy*
6:30 SOCIETY FOR ASIAN AND COMPARATIVE PHILOSOPHY, *Monarch*
6:30 NORTH AMERICAN SOCIETY FOR SOCIAL PHILOSOPHY, *Province*

9:00 RECEPTION, *Imperial Ballroom*

WEDNESDAY, DECEMBER 30

9:00 A.M.–12:00 NOON REGISTRATION, *Rear Main Lobby*
9:00 A.M.–2:00 P.M. BOOK DISPLAYS, *Sheraton Exhibit Center*

PLACEMENT INFORMATION

For Interviewers, 8:00 A.M.–12:00 NOON, *Versailles Terrace*
For Candidates, 9:00 A.M.–12 NOON, *Princess Ballroom*

MORNING

9:00–11:00 Session IV

A. Symposium: Feminist Perspectives on Individual Choice *Royal Ballroom A*

Chair: Wendy Donner
Speaker: Diana T. Meyers
Commentator: Kathryn Pyne Addelson

B. Symposium: Mind Matters, *Royal Ballroom B*
Chair: Terry Horgan
Speakers: Ernest LePore and Barry Loewer
Commentator: Jerry Fodor

C. Invited Paper: American Philosophy (Mead), *Imperial Ballroom B*
Chair: John McDermott
Hans Joas, Meadian Pragmatism and Social Theory
Commentator: James Campbell

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Wednesday, December 30, Morning, Session IV

D. Colloquium: Political and Social Philosophy, *Versailles Ballroom*
   1. *Chair:* Gerard Elfstrom
      Reidar K. Lie, Dworkin on External Preferences
      Commentator: Geoffrey Sayre McCord
   2. *Chair:* Roger S. Gottlieb
      Douglas Ehring, Cohen, Exploitation, and Theft
      Commentator: Scott Arnold

E. Colloquium: Philosophy of Language, *Georgian Ballroom A*
   1. *Chair:* James Higginbotham
      Lawrence Roberts, Quantifier Order, Reflexive Pronouns, and Quasi-indexicals
      Commentator: William J. Rapaport
   2. *Chair:* Margaret Crouch
      Roy A. Sorensen, Vagueness Implies Cognitivism
      Commentator: Kelly Alberts

F. Colloquium: Ancient Philosophy, *Georgian Ballroom B*
   1. *Chair:* Scott Austin
      Thomas C. Brickhouse and Nicholas D. Smith, Socrates's Elenctic Mission
      Commentator: Richard Kraut
   2. *Chair:* Daniel C. Shartin
      Robin Smith, The Educative Function of *Metaphysics* Gamma
      Commentator: Tom Upton

G. Colloquium: History of Philosophy, *Regency Ballroom*
   1. *Chair:* Mark Bernstein
      David G. Stern, Wittgenstein's Phenomenology and the "Remarks on Logical Form"
      Commentator: Grant Luckhardt
   2. *Chair:* Nicholas Capaldi
      Thomas C. Vinci, Distinct Ideas and the Benevolence Principle in Descartes
      Commentator: Susan Bordo

H. Special Symposium Arranged by the APA Committee on Precollege Philosophy, *Diplomat*
   Topic: Teaching Thinking Skills: The North Carolina Model
   Chair: Gareth B. Matthews
   Speaker: Anthony G. Rud, Jr.
   Commentators: Harvey Siegel and Paul Wagner

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Group Meetings (See Appendix for Programs)

9:00 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Commonwealth*

9:00 NORTH AMERICAN NIETZSCHE SOCIETY, *Senate*

11:15 A.M.–12:15 P.M. BUSINESS MEETING (Second Session, if needed) *Imperial Ballroom A*

AFTERNOON

1:30–4:30 Session V

A. Symposium: Enlightenment and Rationality *Royal Ballroom A*

Chair: Raymond Geuss
Speakers: Fred Dallmayr and Axel Honneth
Commentator: Tom Rockmore

B. Symposium: Aristotle's *Metaphysics, Royal Ballroom B*
Chair: John Anton
Speakers: Edward Halper and Charlotte Witt
Commentator: Sarah W. Broadie

C. Invited Papers: *Imperial Ballroom B*
1. Philosophy of Religion (Occasionalism)
Chair: Daisie Radner
Philip Quinn, Divine Conservation, Secondary Causes, and Occasionalism
Commentator: John Fischer
2. Oriental Philosophy (Confucian Ethics)
Chair: Ellen Chen
Antonio Cua, The Concept of *Li* (Proper Behavior) in Confucian Moral Theory
Commentator: Lik Kuen Tong

D. Colloquium: Philosophy of Mind, *Versailles Ballroom*
1. *Chair:* Robert van Gulick
Susan Brison, On Fodor's A Priori Defense of the Language-of-thought Hypothesis
Commentator: Ivan Fox
2. *Chair:* Andrew C. Ward
William Bechtel, Connectionism and Rules and Representation Systems: Are They Compatible?
Commentator: Dan Lloyd
3. *Chair:* Brian McLaughlin
Rosalind Hursthouse, A New Argument against the Causal Theory of Intentional Actions
Commentator: Richard Liebendorfer

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Wednesday, December 30, Afternoon, Session V

E. Colloquium: Moral Psychology, *Georgian Ballroom A*
   1. *Chair:* Jacob Owensby
      Richard F. Galvin, Does Kant's Psychology of Morality Need Basic Revision?
      Commentator: Roger Sullivan
   2. *Chair:* Gary Watson
      Judith M. Hill, Reason and Motivation
      Commentator: Christine Korsgaard
   3. *Chair:* O. Harvey Green
      John Christman, In Defense of Autonomy
      Commentator: Joel Marks

F. Colloquium: Applied Ethics, *Regency Ballroom*
   1. *Chair:* Raymond Belliotti
      Sara Ann Ketchum, Selling Babies and Selling Bodies: Surrogate Motherhood and the Problem of Commodification
      Commentator: Heidi Malm
   2. *Chair:* Raphael Sassower
      Patricia Illingworth, The Friendship Model of Physician/Patient Autonomy
      Commentator: David James
   3. *Chair:* Jonathan Schonsheck
      Richard B. Miller, A Meritocratic Argument for Preferential Treatment
      Commentator: Judith Decew

G. Discussion of Recent Notable Books: *Georgian Ballroom B*
   1. *Foundations of Space-Time Theories* by Michael Friedman
      Chair: Adolf Grünbaum
      Discussant: Lawrence Sklar
      Respondent: Michael Friedman
   2. *Kant and the Claims of Knowledge* by Paul Guyer (Kant Bicentennial)
      Chair: Francis Wilson
      Discussant: Patricia Kitcher
      Respondent: Paul Guyer

1:30 SOCIETY FOR WOMEN IN PHILOSOPHY, *Regency Foyer*
1:30 EXECUTIVE COMMITTEE MEETING, *St. James B*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

PROGRAM PARTICIPANTS
EIGHTY-FOURTH ANNUAL MEETING

ACKERMAN, Bruce A., Columbia University Law School I-G
ADDELSON, Kathryn Pyne, Smith College, Massachusetts IV-A
ADLER, Jonathan, Brooklyn College, City University of New York III-A
ALBERTS, Kelly, University of Wyoming IV-E
ALLAN, George, Dickinson College, Pennsylvania III-E
ANNAS, Julia, University of Arizona I-C
ANNIS, David B., Ball State University, Indiana II-F
ANTON, John, University of South Florida V-B
ARNOLD, Scott, University of Alabama/Birmingham IV-D
ASHBAUGH, Anne, Colgate University III-F
AUSTIN, Scott, Boston University IV-F
BAILEY, George, East Carolina University, North Carolina II-G
BARBER, Richard, University of Louisville III-B
BAXTER, Donald, Princeton University I-E
BEALER, George, Reed College, Oregon I-A
BECHTEL, William, Georgia State University V-D
BEER, Michelle, Florida International University/Miami I-E
BELLIOTTI, Raymond, State University of New York/Fredonia V-F
BENNETT, Jonathan, Syracuse University Presidential Address
BERNSTEIN, Mark, University of Texas/San Antonio IV-G
BERNSTEIN, Richard J., Haverford College, Pennsylvania III-C
Introduction, Presidential Address
BOISVERT, Raymond, Siena College, New York II-D
BORDO, Susan, Le Moyne College, New York IV-G
BOXILL, Bernard, University of North Carolina/Chapel Hill III-H
BRICKHOUSE, Thomas C., Lynchburg College, Virginia IV-F
BRISON, Susan, Dartmouth College, New Hampshire V-D
BROADIE, Sarah W., Yale University V-B
BROCK, Dan, Brown University III-B
CAHN, Steven M., Graduate Center, City University of New York II-H
CAMPBELL, James, University of Toledo IV-C
CAPALDI, Nicholas, Queens College, City University of New York IV-G
CHEN, Ellen, St. John's University, New York V-C
CHENEY, James, University of Wisconsin/Waukesha I-I
CHISHOLM, Roderick, Brown University I-H
CHRISTMAN, John, Virginia Polytechnic Institute V-E
CHURCHLAND, Patricia Smith, University of California/La Jolla I-A
CLARKE, Bowman, University of Georgia III-E
CONEE, Earl, University of Rochester II-E
CRAIG, William L., Westmont College, California I-F
CROUCH, Margaret, Eastern Michigan University IV-E
CUA, Antonio, Catholic University of America V-C

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CULVER, Catherine M., Florida State University III-F
CURRY, Blanche, Shaw University III-H
CUSHING, James, University of Notre Dame I-C
DALLMAYR, Fred, University of Notre Dame V-A
DANIELS, Norman, Tufts University III-B
DE NYS, Martin, George Mason University II-B
DECEW, Judith, Massachusetts Institute of Technology V-F
DIAMOND, Irene, University of Oregon I-I
DONNER, Wendy, University of Regina, Canada IV-A
DRAPER, Paul, University of Wisconsin/Oshkosh I-F
DRYDYK, Jay, University of Toronto, Canada II-B
DURAN, Jane, University of California/Santa Barbara I-E
EARMAN, John, University of Pittsburgh I-F
EHRING, Douglas, Southern Methodist University IV-D
ELFSTROM, Gerard, Agnes Scott College, Georgia IV-D
FERREIRA, Jaimie, University of Virginia I-F
FISCHER, John, Yale University V-C
FLAY, Joseph, Pennsylvania State University/University Park II-B
FODOR, Jerry, Graduate Center, City University of New York IV-B
FORD, Lewis S., Old Dominion University III-E
FOX, Ivan, Yale University V-D
FRIEDMAN, Michael, University of Illinois/Chicago V-G
FUCHS, Allen, College of William and Mary, Virginia II-E
GALVIN, Richard F., Texas Christian University V-E
GARBER, Daniel, University of Chicago II-A
GARLAND, William, University of the South III-E
GARRETT, Roland, Montclair State College, New Jersey II-D
GEUSS, Raymond, Columbia University V-A
GOLDING, Martin, Duke University I-G
GOLDMAN, Alan, University of Miami, Florida II-C
GOLDMAN, Alvin I., University of Arizona I-A
GOTTLIEB, Roger S., Worcester Polytechnic Institute IV-D
GOUINLOCK, James, Emory University II-D
GRANDY, Richard, Rice University I-D
GREEN, O. Harvey, Tulane University V-E
GRÜNBAUM, Adolf, University of Pittsburgh V-G
GUYER, Paul, University of Pennsylvania V-G
HALLER, Rudolf, Graz University, Austria I-H
HALPER, Edward, University of Georgia V-B
HAMPTON, Cynthia, Ohio University III-F
HAND, Michael, University of Oklahoma III-D
HANSEN, Chad D., University of Vermont II-H
HANSON, Karen, Indiana University/Bloomington II-H
HARPER, William, University of Western Ontario, Canada I-D
HARRIS, James, College of William and Mary, Virginia I-F
HART, Richard E., Bloomfield College, New Jersey II-G

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

HEELAN, Patrick, State University of New York/Stony Brook I-E
HEIL, John, Davidson College, North Carolina I-C
HIGGINBOTHAM, James, Massachusetts Institute of Technology IV-E
HILL, Judith M., George Mason University V-E
HONNETH, Axel, Frankfurt University, West Germany V-A
HOOKER, Michael, University of Maryland/Baltimore II-H
HORGAN, Terry, Memphis State University, Tennessee IV-B
HURSTHOUSE, Rosalind, Open University, England V-D
ILLINGWORTH, Patricia, McGill University V-F
JAGGAR, Alison M., University of Cincinnati I-I
JAMES, David, Longwood College, Virginia V-F
JOAS, Hans, Erlangen-Nürenberg University, Sociology IV-C
JOHNSEN, Bredo C., University of Houston III-A
JOY, Lynn, Vanderbilt University II-A
KAMITZ, Reinhard, Graz University, Austria I-H
KASACHKOFF, Tziporah, Borough of Manhattan Community College III-I
KETCHUM, Sara Ann, Harvard Law School V-F
KING, Ynestra, Institute for Social Ecology, Vermont I-I
KITCHER, Patricia, University of California/San Diego V-G
KOBES, Bernard, Arizona State University I-D
KORSGAARD, Christine, University of Chicago V-E
KRAUT, Richard, University of Illinois/Chicago IV-F
KYBURG, Henry E., Jr., University of Rochester I-D
LARRABEE, Mary Jeanne, DePaul University I-E
LEE, Sander H., Keene State College, New Hampshire II-F
LEGENHAUSEN, Gary, Texas Southern University I-F
LEHRER, Keith, University of Arizona I-H
LEPORE, Ernest, Rutgers University/New Brunswick IV-B
LIE, Reidar K., East Carolina University, North Carolina IV-D
LIEBENDORFER, Richard, University of Louisville V-D
LINDERMAYER, Eric, Suffolk Community College, SUNY III-I
LIPSON, Morrie, Vancouver, British Columbia, Canada III-D
LLOYD, Dan, Trinity College, Connecticut V-D
LOEWER, Barry, University of South Carolina/Columbia IV-B
LOSONSKY, Michael, University of Minnesota/Duluth III-D
LOTT, Tommy, University of Massachusetts/Boston III-G
LUCAS, Billy Joe, Manhattanville College, New York I-F
LUCAS, George R., Clemson University III-E
LUCKHARDT, Grant, Georgia State University IV-G
MACHAMER, Peter, University of Pittsburgh I-C
MACINTYRE, Alasdair, Vanderbilt University III-C
MALM, Heidi, University of Nebraska/Lincoln V-F
MARKS, Joel, University of New Haven V-E
MATTHEWS, Gareth B., University of Massachusetts/Amherst IV-H
MAYO, Deborah, Virginia Polytechnic Institute I-D
MCBRIDE, William, Purdue University/West Lafayette II-G

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

McCAULEY, Robert, Emory University I-A
McCONNELL, Terrance, University of North Carolina/Greensboro II-E
McCORD, Geoffrey Sayre, University of North Carolina/Durham IV-D
McCORMICK, Peter, University of Ottawa, Canada I-E
McDERMOTT, John, Texas A & M University IV-C
McDOWELL, John, University of Pittsburgh II-C
McGRADE, Arthur S., University of Connecticut/Storrs III-G
McINTYRE, Ronald, California State University/Northridge I-B
McLAUGHLIN, Brian, Rutgers University/New Brunswick V-D
McMULLIN, Ernan, University of Notre Dame I-F
McPHERRAN, Mark, University of Maine III-F
MENDELL, Mark, University of Pennsylvania II-D
MEYERS, Diana T., University of Connecticut/Storrs IV-A
MICHELMAN, Frank I., Harvard Law School I-G
MILLER, Clyde Lee, State University of New York/Stony Brook III-G
MILLER, Richard B., East Carolina University, North Carolina V-F
MOHANTY, Jitendra N., Temple University I-B
MORSCHER, Edgar, University of Salzburg, Austria I-H
MORRIS, William E., University of Cincinnati III-D
MOSER, Paul, Loyola University of Chicago III-A
O'CONNOR, David, University of Notre Dame II-F
O'NEILL, Eileen, Queens College, City University of New York II-A
OWENSBY, Jacob, Jacksonville University V-E
PLECHA, James, Oak Park, Illinois I-E
POJMAN, Louis, University of Mississippi I-F
POPKIN, Richard H., Washington University, St. Louis I-C
POWELL, C. Thomas, Guilford College, North Carolina III-D
PURNELL, Fred, Queens College, City University of New York III-G
QUINN, Philip, University of Notre Dame V-C
RADNER, Daisie, State University of New York/Buffalo V-C
RAPAPORT, William J., State University of New York/Buffalo IV-E
RECK, Andrew, Tulane University I-G
RHODES, Rosamond, Bronx, New York III-G
ROBERTS, Lawrence, State University of New York/Binghamton IV-E
ROCKMORE, Tom, Duquesne University V-A
ROSSMAN, Neil, LaGuardia Community College, CUNY III-I
RUD, Anthony G., Jr., Western Carolina University, North Carolina IV-H
SANKOWSKI, Edward, University of Oklahoma II-F
SASSOWER, Raphael, University of Colorado/Colorado Springs V-F
SAVITT, Steven F., University of British Columbia, Canada III-D
SCANLON, Thomas, Harvard University II-C
SCHOEMAN, Ferdinand, University of South Carolina/Columbia II-E
SCHONSHECK, Jonathan, Le Moyne College, New York V-F
SCHWEBER, Sylvan S., Brandeis University I-C
SEIDENFELD, Teddy, Carnegie-Mellon University I-D
SHAPIRO, Gary, University of Kansas II-G
SHARTIN, Daniel C., Clark University IV-F

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SHERBURNE, Donald, Vanderbilt University III-E
SHEROVER, Charles, Hunter College, City University of New York I-E
SHRADER-FRECHETTE, Kristin, University of South Florida I-E
SIEGEL, Harvey, University of Miami, Florida IV-H
SINGER, Beth, Brooklyn College, City University of New York II-D
SKLAR, Lawrence, University of Michigan, Ann Arbor V-G
SLAUGHTER, Thomas, Bentley College, Massachusetts III-H
SMITH, Nicholas D., Virginia Polytechnic Institute IV-F
SMITH, P. Christopher, University of Lowell III-F
SMITH, Robin, Kansas State University IV-F
SOKOLOWSKI, Robert, Catholic University of America I-B
SORENSEN, Roy A., University of Delaware IV-E
STERN, David G., University of California/Berkeley IV-G
STEWART, Rod, Austin College, Texas I-D
SULLIVAN, Roger, University of South Carolina/Columbia V-E
TALBOTT, W. J., Seattle, Washington I-D
TEJERA, Victorino, State University of New York/Stony Brook II-D
TONG, Lik Kuen, Fairfield University V-C
TRIANOWSKI, Gregory, University of North Carolina/Chapel Hill II-E
UPTON, Tom, Gannon University IV-F
VALLENTYNE, Peter, University of Western Ontario, Canada II-E
VAN DER LINDEN, Harry, Colgate University III-G
VAN GULICK, Robert, Syracuse University V-D
VERENE, Donald P., Emory University III-C
VICKERS, John M., Claremont Graduate School I-D
VINCI, Thomas C., Dalhousie University, Canada IV-G
VOGEL, Jonathan, Amherst College, Massachusetts III-D
WAGNER, Paul, University of Houston/Clear Lake IV-H
WARD, Andrew C., University of Minnesota/Minneapolis V-D
WARREN, Karen J., Macalester College, Minnesota I-I
WATSON, Gary, University of California/Irvine V-E
WATSON, Stephen, University of Notre Dame II-G
WEINGARTNER, Paul, University of Salzburg, Austria I-H
WELTON, Donn, State University of New York/Stony Brook I-B
WHITE, David A., DePaul University III-F
WHITING, Jennifer, University of Pittsburgh II-F
WILES, Ann, James Madison University II-F
WILLIAMS, Robert R., Hiram College, Ohio II-G
WILSON, Francis, Bucknell University V-G
WITT, Charlotte, University of New Hampshire V-B

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

APPENDIX: PROGRAMS OF GROUP MEETINGS

SUNDAY, DECEMBER 27

I. EVENING (8:00–11:00)

1. 8:00 SOCIETY OF CHRISTIAN PHILOSOPHERS, *Royal Ballroom B*
   Topic: Religion and Technology
   Chair: Clement J. Dore
   I. John Post, On Reenchanting the World
   II. Submitted Essay
   Commentator: Frederick Ferré

2. 8:00 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Versailles Ballroom*
   Topic: Workshop on Philosophy and the Holocaust
   Chair: Sander H. Lee
   I. Berel Lang, Genocide and Kant's Enlightenment
   II. Ronald Aronson, Responsibility for the Holocaust
   III. Business Meeting

3. 8:00 SOCIETY FOR CLASSICAL REALISM, *Senate*
   Topic: An interdisciplinary Approach to the Nature of Truth
   Chair: Mary Carmen Rose
   I. Gertrude R. Schmeidler, Truth: Does It Keep Changing?
   II. Christopher Russell, Truth and Evolution
   Commentator: James Kelly

4. 8:00 THE SOCIETY OF PHILOSOPHERS AT WORK IN THE WORLD *Regency Foyer*
   Co-chairs: Jean Campbell and Nance Cunningham Butler
   I. *Topic:* Philosophers' Perspectives on Project Management in the Trenches
   II. Business Meeting

MONDAY, DECEMBER 28

II. MORNING (9:00–11:00)

1. 9:00 NORTH AMERICAN NIETZSCHE SOCIETY, *Royal Ballroom B*
   Chair: R. J. Hollingdale
   I. Robin Small, Absolute Becoming and Absolute Necessity
   Commentator: John Richardson
   II. Richard White, Nietzsche contra Kant and the Problem of Autonomy
   Commentator: Mark Fowler
   III. Daniel Conway, A Moral Ideal for Everyone and No One: Nietzsche's Oblique Promotion of Virtue
   Commentator: Lester Hunt
   IV. Business Meeting (Executive Secretary: Bernd Magnus)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

2. 9:00 RADICAL PHILOSOPHY ASSOCIATION, *Versailles Ballroom*
   Topic: Education and Critical Consciousness
   Chair: Leonard Harris
   I. Gene James, Paolo Freire's Philosophy of Education
   II. Paula Rothenberg, Teaching Gender, Race, and Class
   Commentator: Otto Begus

3. 9:00 SOCIETY FOR WOMEN IN PHILOSOPHY, *Georgian Ballroom A*
   I. Business Meeting
      Chair: Elizabeth Potter
   II. In Honor of Distinguished Women Philosophers
      Chair: Ernan McMullin
      Honoree: Elizabeth Flower
      a. *Speaker:* Kathleen Wright
         Commentator: Anne Ashbaugh
      b. *Speaker:* Marilyn Friedman
         Commentator: Virginia Held

4. 9:00 ASSOCIATION FOR SYMBOLIC LOGIC, *Georgian Ballroom B*

5. 9:00 INTERNATIONAL PHILOSOPHERS FOR THE PREVENTION OF NUCLEAR OMNICIDE and CONCERNED PHILOSOPHERS FOR PEACE *Regency Ballroom*
   Topic: Philosophical Perspectives on the Quest for Peace between East and West
   Chair: Howard Friedman
   I. Lyle V. Anderson, Preventing Accidental Nuclear War
   II. Barry L. Gan, Pacifism: A New Perspective for the Super Powers
   III. George H. Hampsch, Peace, Coexistence, and the Ideological Struggle

6. 9:00 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Regency Foyer*
   Topic: Workshop on Philosophy and the Holocaust
   Chair: John Abbarno
   I. Sander H. Lee, The Law and Morality in Crime Trials
   II. Gray Cox, The Holocaust from a Gandhian Perspective

7. 9:00 INTERNATIONAL THOMAS AQUINAS SOCIETY, *Commonwealth*
   Chair: Mary Rose Barral
   Mark Jordan, The Evidence of the Transcendentals and the Place of Beauty

8. 9:00 SOCIETY FOR THE STUDY OF PROCESS PHILOSOPHIES, *Senate*
   Topic: Allan's *Importance of the Past*
   I. Charles Sherover, The Future of the Past
   II. Joseph Grange, Norms, Transition, and Value
   Commentator: George Allan

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Monday, December 28, Morning, II

9. 9:00 SOCIETY FOR PHILOSOPHY AND TECHNOLOGY, *Dominion A*
   Chair: Paul T. Durbin
   Don Ihde, Technology and the Life-World
   Commentator: Robert Ackerman

10. 9:00 NORTH AMERICAN SOCIETY FOR SOCIAL PHILOSOPHY, *Dominion B*
   Topic: The American Constitution: Prospects and Retrospect
   Chair: Blanche Curry
   I. James Gould, Privacy versus Liberty in *Roe vs. Wade*
      Commentator: Eugene Kelly
   II. Sterling Harwood, Should Judges Follow Precedent?
      Commentator: James P. Friel
   III. Christopher P. Gray, On Amending the United States Constitution

11. 9:00 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Province*
   Topic: Symposium on Contemporary Antifoundationalism and the Prospects for Systematic Philosophy: Analytic Perspectives
   Chair: Kevin Clark
   I. Bill Martin, Why Go Transcendental when You Can Simply Triangulate? (Why Davidson Does Not Need Transcendental Arguments)
   II. William Maker, Davidson's Transcendental Arguments
   Commentator: Frank Farrell

12. 9:00 BERTRAND RUSSELL SOCIETY, *Embassy*
   I. Jan Dejnozka, Russell's Robust Sense of Reality: A Reply to Butchvarov
   II. Russell Wahl, On What Is Denoted

13. 9.00 APA COMMITTEE ON INTERNATIONAL COOPERATION, SOCIEDAD FILOSOFICA IBEROAMERICANA, and SOCIETY FOR IBERIAN AND LATIN AMERICAN THOUGHT, *Consulate*
   Topic: Philosophy in Latin America Today
   Chair: Ernest Sosa
   I. Ezequiel de Olaso, Some Considerations on Latin American Philosophy Today
   II. León Olivé, Cultural Relativism and Conceptual Relativism
   III. Enrique Villanueva, Relativism and Philosophy in Latin America Today
   IV. Jorge J. E. Gracia, Philosophy and Cultural Identity

14. 9:00 ASSOCIATION FOR PHILOSOPHY OF THE UNCONSCIOUS, *Diplomat*

III. EARLY AFTERNOON (11:15–1:15)

1. 11:15 SOCIETY FOR THE ADVANCEMENT OF AMERICAN PHILOSOPHY *Versailles Ballroom*
   Topic: Metaphysics after Pragmatism
   Chair: John Lachs
   Speaker: David Weissman
   Commentator: John Ryder

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

2. 11:15 AMERICAN SOCIETY FOR VALUE INQUIRY, *Regency Ballroom*
   Chair: Lisa Newton
   Presidential Address: Patricia Werhane, Moral Realism
   Business Meeting

3. 11:15 SARTRE CIRCLE, *Regency Foyer*
   Topic: Sartre on the Emotions
   Chair: Phyllis S. Morris
   Speakers: Peter Caws, Jennifer Church, and Robert C. Solomon
   Commentators: Frithjof H. Bergmann

4. 11:15 INTERNATIONAL SOCIETY FOR NEOPLATONIC STUDIES, *Diplomat*
   Topic: The Place of the Individual in Neoplatonic Thought
   Chair: William J. Carroll
   I. John Jones, Whither Timothy: Individuality and the Quest for *Henosis* in Pseudo-Dionysius the Areopagite
   II. Arthur Madigan, Plotinus on Personal Individuality: Early/Late

5. 11:15 ASSOCIATION FOR PHILOSOPHY AND LIBERATION, *Commonwealth*
   Chair: Vincent Punzo
   James L. Marsh, The Bishops' Pastoral on Economic Justice: A Neo-Marxist Interpretation
   Commentator: Martin De Nys

6. 11:15 JASPERS SOCIETY OF NORTH AMERICA, *Senate*
   Chair: Robert S. Corrington
   I. Gregory J. Walters, Karl Jaspers on the Role of "Conversion" in the Nuclear Age
      Commentator: Michael Quirk
   II. William Kluback, Idolatry in the Political Thought of Karl Jaspers
      Commentator: Harold Oliver

7. 11:15 CONCERNED PHILOSOPHERS FOR PEACE, *Dominion A*
   Chair: Barry L. Gan

8. 11:15 MOORE SOCIETY, *Embassy*
   Topic: The Early Moore
   Chair: Dennis A. Rohatyn
   Tom Regan, Moore: The Liberator
   Commentators: Abraham Edel, Jaakko Hintikka

9. 11:15 WORLD PHENOMENOLOGY INSTITUTE, *Monarch*
   Topic: Social Justice and the Moral Sense
   I. William McBride, Social Justice on Trial: The Verdict of History
   II. Anna-Teresa Tymieniecka, Is Social Justice Possible? The Return to Moral Sense

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

MONDAY, DECEMBER 28

IV. AFTERNOON (2:00–5:00)

1. 2:00 ASSOCIATION FOR SYMBOLIC LOGIC, *Georgian Ballroom B*

2. 2:00 AMERICAN ASSOCIATION OF PHILOSOPHY TEACHERS, *Diplomat*
   Chair: Rosalind Ekman Ladd
   Alasdair MacIntyre, Education as Initiation into Conflict
   Commentator: Donald Reed

3. 2:00 AMERICAN ASSOCIATION FOR THE PHILOSOPHIC STUDY OF SOCIETY and THE INSTITUTE FOR HUMANE STUDIES, *Senate*
   Chair: Douglas B. Rasmussen
   I. J. Roger Lee, Friendship and the Self
   II. Neera Badhwar, Utilitarianism and Friendship
   III. Lester Hunt, Friendly Relations and the Basis of Character

4. 2:00 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING *Consulate*
   Chair: Mark L. Weinstein
   I. Bertram Bandman, The Role of Belief in Critical Thinking
      Commentator: Donald Hatcher
   II. Gordon E. Whitney, Analysis of a Dialogue on Determinism and Free Will
      Commentator: Philip A. Pecorino

V. LATE AFTERNOON (5:15–7:15)

1. 5:15 RADICAL PHILOSOPHY ASSOCIATION, *Versailles Ballroom*
   Topic: Equality and Difference: Issues in Race and Gender
   Chair: Clifford Durand
   I. Sara Begus, Portia's Dilemma: Feminist Jurisprudence and the Politics of Identity
   II. Lucius Outlaw, The Politics of Equality: A Double-edged Sword
   Commentator: Iris Young

2. 5:15 SOCIETY FOR WOMEN IN PHILOSOPHY, *Regency Ballroom*
   Merrill Hintikka Symposium
   Topic: Women's Education
   Panel: Kathryn Pyne Addelson, Jane Martin, Elizabeth Spelman, and C. J. B. MacMillan
   Moderator: Jaakko Hintikka

3. 5:15 PEIRCE SOCIETY, *Regency Foyer*
   Chair: John Lachs
   I. H. P. Joswick, Peirce's Semiotic: Logic as a Theory of Reasoning
      Commentator: Robert W. Burch
   II. Peirce Society Essay Contest: Award Paper

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

4. 5:15 SOCIETY FOR THE PHILOSOPHICAL STUDY OF THE PARANORMAL *Diplomat*

Topic: How Para Are Parapsychological Powers?
Chair: Lewis Foster
I. Hoyt Edge, Parapsychology or Anomalies Research?
II. Frank B. Dilley, Mind-Body Interaction and Parapsychological Powers

5. 5:15 INTERNATIONAL SOCIETY FOR CHINESE PHILOSOPHY *Commonwealth*

Topic: Habermas and Chinese Philosophy
Chair: Chung-ying Cheng
I. Chung-ying Cheng, On Communicative Rationality from the Confucian Point of View
II. Yu hui Park, Habermas and Confucius
III. Lauren Pfister, Habermas and the Concept of Great Unity

6. 5:15 GANDHI-KING SOCIETY, *Senate*
Chair: Newton Garver
Discussion: Can There Be Justice without Peace?

7. 5:15 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY, *Consulate*
I. Peter King, Peter Abelard's Philosophy of Mind
II. Allan Bäck, Anselm on Perfect Islands

8. 5:15 SOCIETY FOR THE STUDY OF ETHICS AND ANIMALS, *Dominion A*
Chair: Harlan B. Miller
I. Evelyn B. Pluhar, Speciesism: A Form of Bigotry or a Justified View?
Commentator: Steve F. Sapontzis
II. Roberta Kalechofsky, Metaphors of Nature: Pornography and Vivisection
Commentator: Laura Westra

9. 5:15 SOCIETY FOR CLASSICAL REALISM, *Embassy*
Chair: Mary Carmen Rose
I. Earle Coleman, Beauty and the "I" of Buber's Beholder
II. Xavier Monasterio, Dramatic Realism

10. 5:15 SOCIETY FOR THE PHILOSOPHICAL STUDY OF MARXISM, *Monarch*
Topic: The U.S. Constitution and Historical Materialism
Chair: John Neumeier
I. Terry di Filippo, On the Importance of Enlightenment Romanist Legal History for Marxist Interpretation of the U.S. Constitution
II. Bertell Ollman, Founding Finaglers: Toward a Marxist Interpretation of the U.S. Constitution
III. John Ryder, Private Property and the U.S. Constitution

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Monday, December 28, Late Afternoon, V

11. 5:15 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Province*
Topic: Symposium on Contemporary Antifoundationism and the Prospects for Systematic Philosophy: Analytic Prospectives
Chair: Robert Berman
I. Kenley R. Dove, Minding Our Language
II. Joseph C. Pitt, Sellarsian Antifoundationalism
Commentator: George Lucas

12. 5:15 SOCIETY FOR BUSINESS ETHICS, *Dominion B*
Topic: Corporate Ethics
Chair: Michael Hoffman
I. R. Edward Freeman, Corporations and Capitalism: A Stakeholder Approach
II. Thomas Donaldson, Economic Injustice and Corporate Takeovers

13. 5:15 ASSOCIATION OF PHILOSOPHY JOURNAL EDITORS *Royal Ballroom B*
Topic: What Journals Owe Philosophers and Vice Versa.
Chair: Arnold Wilson

VI. EVENING (7:30–10:30)

1. 7:30 NORTH AMERICAN KANT SOCIETY, *Versailles Ballroom*
Speaker: Hilary Putnam, Realism and the Life of Reason
Commentator: Gordon Brittan

2. 7:30 AMERICAN SOCIETY FOR VALUE INQUIRY, *Regency Ballroom*
Chair: Sander H. Lee
Robert Ginsberg, The Value of Philosophy: A Dialogue

3. 7:30 SANTAYANA SOCIETY, *Regency Foyer*
Chair: Paul G. Kuntz
John Lachs, The Enduring Value of Santayana's Philosophy
Commentator: Herman J. Saatkamp

4. 7:30 LEIBNIZ SOCIETY OF AMERICA, *Senate*
Chair: Margaret Wilson
I. François Duchesneau, Leibniz on Architectonic Principles and Analysis
II. Business Meeting
Presiding: Margaret Wilson, President

5. 7:30 SOCIETY FOR IBERIAN AND LATIN AMERICAN THOUGHT, *Consulate*
Chair: Jorge J. E. Gracia
I. Anton Donoso and Roberto J. Vichot, Dewey, Ortega, and Pragmatism
II. Ezequiel de Olaso, Ideas for the Critical Edition of Ortega's Works

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

6. 7:30 ASSOCIATION FOR SYMBOLIC LOGIC, *Dominion A*
   Council Meeting
   Chair: Michael Morley

7. 7:30 SOCIETY FOR PHILOSOPHY AND PUBLIC AFFAIRS, *Dominion B*
   Chair: Diana T. Meyers
   Christina Hoff Sommers, The Philosopher's War against the Family
   Commentators: Alison M. Jaggar and Kathryn Jackson

8. 7:30 PERSONALIST DISCUSSION GROUP, *Embassy*
   Chair: Thomas O. Buford
   Eugene Fontinelle, Personal Immortality: Life-enhancing or Life-eroding?

9. 7:30 PHILOSOPHERS FOR SOCIAL RESPONSIBILITY, *Monarch*
   Topic: Pornography

10. 7:30 WORLD PHENOMENOLOGY INSTITUTE, *Province*
    Topic: Social Justice and Moral Sense
    I. David Doyle, The Justice of Mercy
    II. William Hamrick, Rules of Law and Substantive Justice

11. 7:30 SOCIETY FOR ANCIENT GREEK PHILOSOPHY, *Royal Ballroom B*
    Topic: Aristotle's Ethics
    Chair: Anthony Preus
    I. Robert Bolton, Aristotle's Ethics
    II. Daniel Devereux, Theoria and Praxis in the *Eudemian* and *Nicomachean Ethics*

TUESDAY, DECEMBER 29

VII. MORNING (9:00–11:00)

1. 9:00 ASSOCIATION FOR SYMBOLIC LOGIC, *Georgian Ballroom B*

2. 9:00 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Diplomat*
   Symposium: Apropos the Bicentennial of the U.S. Constitution
   Chair: Deborah Chaffin
   I. Richard Dien Winfield, Constitutionality and Political Foundationalism
   II. Frank Bryce McCluskey and George Teschner, Deconstructing the First Amendment: Philosophical Presuppositions for the City in Speech

3. 9:00 SOCIETY FOR THE PHILOSOPHY OF SEX AND LOVE, *Commonwealth*
   Alan Soble, On the Coherence of Love (Tenth Anniversary Presidential Address)

4. 9:00 SOCIETY FOR THE PHILOSOPHY OF CREATIVITY, *Senate*
   Chair: Paul G. Kuntz
   Topic: Paul Weiss on Creativity
   Speaker: Paul Weiss
   Commentators: Charles Hartshorne, Carl Hausman, William Desmond, and Peter A. Y. Gunter

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Tuesday, December 29, Morning, VII

5. 9:00 SARTRE CIRCLE, *Embassy*
   I. *Chair:* Debra Bergoffen
   Ronald E. Santoni, The Cynicism of Sartre's Bad Faith
   Commentator: Joseph S. Catalano
   II. *Chair:* Philip L. Peterson
   Eleanor Kuykendall, Sartre's Linguistic Phenomenology
   Commentator: Robert V. Stone

VIII. EARLY AFTERNOON (11:15–1:15)

1. 11:15 RADICAL PHILOSOPHY ASSOCIATION, *Versailles Ballroom*
   Topic: Science, Gender, Race, and Class
   Chair: Val Dusek
   I. Anthony Appiah, Social Forces/Natural Kinds
   II. Nancy Holmstrom, Biologism, Essentialism, and Feminism
   Commentators: Anne Fausto-Sterling, Gena Porto, and Lorenzo Simpson

IX. AFTERNOON (2:15–4:15)

1. 2:15 ASSOCIATION FOR SYMBOLIC LOGIC, *Georgian Ballroom B*

2. 2:15 SOREN KIERKEGAARD SOCIETY, *Commonwealth*
   Topic: Kierkegaard and Deconstruction
   Chair: Merold Westphal
   Speakers: John Caputo, Louis Mackey, and Mark Taylor

3. 2:15 CHAIRS' CAUCUS, *Senate*
   Topic: The Next Decade: What Will It Mean for Philosophy Programs?
   Chair: Frank B. Dilley
   Panel: Samuel Gorovitz, David A. Hoekema, and Anita Silvers

4. 2:15 AMERICAN SOCIETY FOR VALUE INQUIRY, *Monarch*
   Chair: Robert Ginsberg
   Charles Hornbeck, The Square of Opposition Vindicated
   Commentators: Stephen F. Barker

X. EVENING (6:30–9:30)

1. 6:30 RADICAL PHILOSOPHY ASSOCIATION and THE SOCIETY FOR THE PHILOSOPHICAL STUDY OF MARXISM, *Versailles Ballroom*
   Topic: Intellectual Foundations of the Nicaraguan Revolution
   Chair: Ed D'Angelo
   Panel: Don Hodges, Martin Vega, and Phil Berryman

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

2. 6:30 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Regency Ballroom*
Symposium: Apropos the Bicentennial of the U.S. Constitution
Topic: The Dialogical Turn in Constitutional Interpretation
Chair: David A. Duquette
I. Bruce A. Ackerman, Does the American Constitution Make Sense?
II. Cass Sunstein, Deliberative Democracy and the American Constitution
III. Frank Michelman
IV. Drucilla Cornell

3. 6:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Regency Foyer*
Topic: Workshop on Philosophy and the Holocaust
Chair: Alan Rosenberg
I. Joan Ringelheim, Politics of Memory and Remembrance
II. Roger S. Gottlieb, Marxism, Feminism, Holocaust
III. Discussion: Establishment of a Philosophical Society on the Holocaust

4. 6:30 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY *Commonwealth*
I. Allan B. Wolter, The Ethics of John Duns Scotus
II. Daniel H. Frank, Humility as a Virtue: A Maimonidean Critique of Aristotle's Ethics

5. 6:30 AMERICAN ASSOCIATION FOR THE PHILOSOPHIC STUDY OF SOCIETY and INSTITUTE FOR HUMANE STUDIES, *Senate*
Topic: Adam Smith's *Theory of Moral Sentiments*
Chair: Douglas den Uyl
I. Charles Griswold, The Impartial Spectator
II. Eugene Heath, The Theory of Moral Sentiments and Spontaneous Order
III. Jeremy Shearmur, From Brother Sense to Brother Man

6. 6:30 JASPERS SOCIETY OF NORTH AMERICA, *Consulate*
Chair: Raymond E. Gogel
Alan M. Olson, *Philosophical Faith and Revelation:* Twenty-five Years Later: A Reassessment
Commentator: George Pepper
Business Meeting

7. 6:30 ASSOCIATION FOR SYMBOLIC LOGIC, *Dominion A*
Council Meeting
Chair: Michael Morley

8. 6:30 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING *Embassy*
Chair: John Hoaglund
I. J. Anthony Blair, Premise Adequacy of Cogent Arguments
Commentator: Dennis A. Rohatyn
II. Open Forum: Priorities for Theory and Practice
III. Business Meeting

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Tuesday, December 29, Evening, X

9. 6:30 SOCIETY FOR ASIAN AND COMPARATIVE PHILOSOPHY, *Monarch*
   Chair and Commentator: David L. Hall
   I. Joel Kupperman, Not in So Many Words: Taoist Strategies of Communication
   II. Lee Siegel, Laughing Matters: Philosophical Humor and Humorous Philosophy in India

10. 6:30 NORTH AMERICAN SOCIETY FOR SOCIAL PHILOSOPHY, *Province*
   Topic: American Social Philosophy
   Chair: Eugene Kelly
   I. Victorino Tejera, Kuklick's Misrepresentation of American Philosophy
   Commentator: Armen Marsoobian
   II. Lucius Outlaw, American Philosophy: The Unfinished American Revolution
   Commentator: Blanche Curry

WEDNESDAY, DECEMBER 30

XI. MORNING (9:00–11:00)

1. 9:00 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Commonwealth*
   Topic: The Range of American Philosophical Practice
   Chair: John A. Loughney
   Speakers: Nicholas Capaldi, Robert C. Neville, John A. Loughney, and Charles Scott

2. 9:00 NORTH AMERICAN NIETZSCHE SOCIETY, *Senate*
   Symposium: Nietzsche's *Nachlass,* Published Corpus and Thought, 1883–89
   In Honor and Memory of Mazzino Montinari
   Chair: Arthur Danto
   Participants: Peter Heller, R. J. Hollingdale, Bernd Magnus, and Richard Schacht

XII. AFTERNOON (1:30–4:30)

1. 1:30 SOCIETY FOR WOMEN IN PHILOSOPHY, *Regency Foyer*
   Topic: Simone de Beauvoir Symposium
   Chair: Joan Ringelheim
   I. Carol Asher, Beauvoir on Death and Dying
   II. Deidre Bair, Beauvoir on Politics and Philosophy

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

GROUP PARTICIPANTS

(See Appendix)

ABBARNO, John, D'Youville College, New York II-6
ACKERMAN, Bruce A., Columbia University Law School X-2
ACKERMAN, Robert, University of Massachusetts/Amherst II-9
ADDELSON, Kathryn Pyne, Smith College, Massachusetts V-2
ALLAN, George, Dickinson College, Pennsylvania II-8
ANDERSON, Lyle V., University of San Diego II-5
APPIAH, Anthony, Cornell University VIII-1
ARONSON, Ronald, Wayne State University, Michigan I-2
ASHBAUGH, Anne, Colgate University II-3
ASHER, Carol, New York XII-1
BÄCK, Allan, Kutztown University of Pennsylvania V-7
BADHWAR, Neera, University of Oklahoma IV-3
BAIR, Deidre, University of Pennsylvania XII-1
BANDMAN, Bertram, Long Island University/Brooklyn IV-4
BARKER, Stephen F., Johns Hopkins University IX-4
BARRAL, Mary Rose, Gannon University II-7
BEGUS, Otto, Morgan State University, Maryland II-2
BEGUS, Sarah, Johns Hopkins University V-1
BERGMANN, Frithjof H., University of Michigan/Ann Arbor III-3
BERGOFFEN, Debra, George Mason University VII-5
BERMAN, Robert, Lang College, New School for Social Research V-11
BERRYMAN, Phil X-1
BLAIR, J. Anthony, University of Windsor, Ontario, Canada X-8
BOLTON, Robert, Rutgers University/New Brunswick VI-11
BRITTAN, Gordon, Montana State University VI-1
BUFORD, Thomas O., Furman University VI-8
BURCH, Robert W., Texas A&M University V-3
BUTLER, Nance Cunningham, Applied Analysis Associates, Denver, Colorado I-4
CAMPBELL, Jean, Brown Brothers Harriman, New York I-4
CAPALDI, Nicholas, Queens College, City University of New York XI-1
CAPUTO, John, Villanova University IX-2
CARROLL, William J., Coppin State College, Maryland III-4
CATALANO, Joseph S., Kean College of New Jersey VII-5
CAWS, Peter, George Washington University III-3
CHAFFIN, Deborah, San Diego State University VII-2
CHENG, Chung-ying, University of Hawaii/Manoa V-5
CHURCH, Jennifer, Vassar College, New York III-3
CLARK, Kevin, Pennsylvania State University II-11
COLEMAN, Earle, Virginia Commonwealth University V-9
CONWAY, Daniel, Stanford University II-1
CORNELL, Drucilla, University of Pennsylvania Law School X-2
CORRINGTON, Robert S., Pennsylvania State University/University Park III-6
COX, Gray, Middle Tennessee University II-6
CURRY, Blanche, Shaw University II-10, X-10
D'ANGELO, Ed, University of Bridgeport X-1
DANTO, Arthur, Columbia University XI-2
DEJNOZKA, Jan, U.S. Naval Academy, Annapolis II-12
DE NYS, Martin, George Mason University III-5

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

DEN UYL, Douglas, Bellarmine College, Kentucky X-5
DE OLASO, Ezequiel, University of Buenos Aires, Argentina II-13, VI-5
DESMOND, William, Loyola College, Baltimore VII-4
DEVEREUX, Daniel, University of Virginia/Charlottesville VI-11
DI FILIPPO, Terry, Rutgers University V-10
DILLEY, Frank B., University of Delaware V-4, IX-3
DONALDSON, Thomas, Loyola University of Chicago V-12
DONOSO, Anton, University of Detroit VI-5
DORE, Clement J., Vanderbilt University I-1
DOVE, Kenley R., State University of New York/Purchase V-11
DOYLE, David, La Trobe University, Australia VI-10
DUCHESNEAU, François, University of Montreal, Canada VI-4
DUQUETTE, David A., St. Norbert College, Wisconsin X-2
DURAND, Clifford, Morgan State University, Maryland V-1
DURBIN, Paul T., University of Delaware II-9
DUSEK, Val, University of New Hampshire VIII-1
EDEL, Abraham, University of Pennsylvania III-8
EDGE, Hoyt, Rollins College, Florida V-4
FARRELL, Frank, State University of New York/Purchase II-11
FAUSTO-STERLING, Anne, Brown University VIII-1
FERRÉ, Frederick, University of Georgia I-1
FLOWER, Elizabeth, University of Pennsylvania II-3
FONTINELLE, Eugene, Queens College, City University of New York VI-8
FOSTER, Lewis, College of William & Mary, Virginia V-4
FOWLER, Mark, College of William & Mary, Virginia II-1
FRANK, Daniel H., University of Kentucky X-4
FREEMAN, R. Edward, University of Virginia V-12
FRIEDMAN, Howard, University of Connecticut/Waterbury II-5
FRIEDMAN, Marilyn, Bowling Green State University, Ohio II-3
FRIEL, James P., State University of New York/Farmingdale II-10
GAN, Barry L., St. Bonaventure University II-5, III-7
GARVER, Newton, State University of New York/Buffalo V-6
GINSBERG, Robert, Pennsylvania State University/Delaware County VI-2, IX-4
GOGEL, Raymond E., Drew University X-6
GOROVITZ, Samuel, Syracuse University IX-3
GOTTLIEB, Roger S., Worcester Polytechnic Institute X-3
GOULD, James, University of South Florida II-10
GRACIA, Jorge J. E., State University of New York/Buffalo II-13, VI-5
GRANGE, Joseph, University of Southern Maine II-8
GRAY, Christopher P., Concordia University, Quebec, Canada II-10
GRISWOLD, Charles, Howard University X-5
GUNTER, Pete A. Y., North Texas State University VII-4
HALL, David L., University of Texas/El Paso X-9
HAMPSCH, George H., College of the Holy Cross, Massachusetts II-5
HAMRICK, William, Southern Illinois University VI-10
HARRIS, Leonard, Morgan State University, Maryland II-2
HARTSHORNE, Charles, University of Texas/Austin VII-4
HARWOOD, Sterling, Cornell University II-10
HATCHER, Donald, Baker University IV-4
HAUSMAN, Carl, Pennsylvania State University/University Park VII-4
HEATH, Eugene, Yale University X-5
HELD, Virginia, Graduate Center, City University of New York II-3

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

HELLER, Peter, State University of New York/Buffalo XI-2
HINTIKKA, Jaakko, Florida State University III-8, V-2
HOAGLUND, John, Christopher Newport College, Virginia X-8
HODGES, Don, University of Florida State/Tallahassee X-1
HOEKEMA, David A., University of Delaware IX-3
HOFFMAN, Michael, Bentley College, Massachusetts V-12
HOLLINGDALE, R. J., London, England II-1, XI-2
HOLMSTROM, Nancy, Rutgers University VIII-1
HORNBECK, Charles, Keene State College, New Hampshire IX-4
HUNT, Lester, University of Wisconsin/Madison II-1, IV-3
IHDE, Don, State University of New York/Stony Brook II-9
JACKSON, Kathryn, Duke University VI-7
JAGGAR, Alison M., University of Cincinnati VI-7
JAMES, Gene, Memphis State University, Tennessee II-2
JONES, John, Marquette University III-4
JORDAN, Mark, University of Notre Dame II-7
JOSWICK, H. P., Swarthmore, Pennsylvania V-3
KALECHOFSKY, Roberta, Micah Publications V-8
KELLY, Eugene, New York Institute of Technology II-10, X-10
KELLY, James, Fordham University I-3
KING, Peter, University of Pittsburgh V-7
KLUBACK, William, Queensborough Community College, CUNY III-6
KUNTZ, Paul G., Emory University VI-3, VII-4
KUPPERMAN, Joel, University of Connecticut/Storrs X-9
KUYKENDALL, Eleanor, State University of New York/New Paltz VII-5
LACHS, John, Vanderbilt University III-1, V-3, VI-3
LADD, Rosalind Ekman, Wheaton College, Massachusetts IV-2
LANG, Berel, State University of New York/Albany I-2, X-2
LEE, J. Roger, California State University/Los Angeles IV-3
LEE, Sander H., Keene State College, New Hampshire I-2, II-6, VI-2
LOUGHNEY, John A., Westfield State College, Massachusetts XI-1
LUCAS, George, Clemson University V-11
MACINTYRE, Alasdair, Vanderbilt University IV-2
MACKEY, Louis, University of Texas/Austin IX-2
MACMILLAN, C. J. B., Florida State University V-2
MADIGAN, Arthur, Boston College III-4
MAGNUS, Bernd, University of California/Riverside II-1, XI-2
MAKER, William, Clemson University II-11
MARSH, James L., Fordham University III-5
MARSOOBIAN, Armen, Southern Connecticut State University X-10
MARTIN, Bill, University of Kansas II-11
MARTIN, Jane, University of Massachusetts/Boston V-2
MCBRIDE, William, Purdue University/West Lafayette III-9
MCCLUSKEY, Frank Bryce, Mercy College, New York VII-2
MCMULLIN, Ernan, Notre Dame University II-3
MEYERS, Diana T., University of Connecticut/Storrs VI-7
MICHELMAN, Frank, Harvard University Law School X-2
MILLER, Harlan B., Virginia Polytechnic Institute & State University V-8
MONASTERIO, Xavier, University of Dayton V-9
MORLEY, Michael, Cornell University VI-6, X-7
MORRIS, Phyllis S., University of Michigan/Dearborn and Ann Arbor III-3
NEUMEIER, John, Empire State College of New York V-10

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NEVILLE, Robert C., Boston University XI-1
NEWTON, Lisa, Fairfield University III-2
OLASO, Ezequiel de, University of Buenos Aires, Argentina II-13, VI-5
OLIVÉ, León, Instituto de Investigaciones Filosoficas, Mexico II-13
OLIVER, Harold, Boston University III-6
OLLMAN, Bertell, New York University V-10
OLSON, Alan M., Boston University X-6
OUTLAW, Lucius, Haverford College, Pennsylvania V-1, X-10
PARK, Yu hui, Smith College, Massachusetts V-5
PECORINO, Philip A., Queensborough Community College, CUNY IV-4
PEPPER, George, Iona College, New York X-6
PETERSON, Philip L., Syracuse University VII-5
PFISTER, Lauren, Baptist College of Hong Kong V-5
PITT, Joseph C., Virginia Polytechnic Institute and State University V-11
PLUHAR, Evelyn B., Pennsylvania State University/Fayette V-8
PORTO, Gena, Stockton State College, New Jersey VIII-1
POST, John, Vanderbilt University I-1
POTTER, Elizabeth, Haverford College, Pennsylvania II-3
PREUS, Anthony, State University of New York/Binghamton VI-11
PUNZO, Vincent, St. Louis University III-5
PUTNAM, Hilary, Harvard University VI-1
QUIRK, Michael, Drew University III-6
RASMUSSEN, Douglas B., St. John's University, New York IV-3
REED, Donald, Western Carolina University, North Carolina IV-2
REGAN, Tom, North Carolina State University III-8
RICHARDSON, John, New York University II-1
RINGELHEIM, Joan, Institute for Research in History, New York City X-3, XII-1
ROHATYN, Dennis A., University of San Diego III-8, X-8
ROSE, Mary Carmen, Goucher College, Maryland I-3, V-9
ROSENBERG, Alan, Queens College, City University of New York X-3
ROTHENBERG, Paula, William Patterson College of New Jersey II-2
RUSSELL, Christopher, Moravian College, Pennsylvania I-3
RYDER, John, State University of New York/Cortland III-1, V-10
SAATKAMP, Herman J., Texas A&M University VI-3
SANTONI, Ronald E., Denison University VII-5
SAPONTZIS, Steve F., California State University/Hayward V-8
SCHACT, Richard, University of Illinois/Urbana-Champaign XI-2
SCHMEIDLER, Gertrude R., City College, City University of New York I-3
SCOTT, Charles, Vanderbilt University XI-1
SHEARMUR, Jeremy, George Mason University X-5
SHEROVER, Charles, Hunter College, City University of New York II-8
SIEGEL, Lee, University of Hawaii X-9
SILVERS, Anita, San Francisco State University, California IX-3
SIMPSON, Lorenzo, University of Richmond VIII-1
SMALL, Robin, Monash University, Australia II-1
SOBLE, Alan, University of New Orleans VII-3
SOLOMON, Robert C., University of Texas/Austin III-3
SOMMERS, Christina Hoff, Clark University VI-7
SOSA, Ernest, Brown University II-13
SPELMAN, Elizabeth, Smith College, Massachusetts V-2
STONE, Robert V., Long Island University/C. W. Post Campus VII-5
SUNSTEIN, Cass, University of Chicago Law School X-2

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

TAYLOR, Mark, Williams College, Massachusetts IX-2
TEJERA, Victorino, State University of New York/Stony Brook X-10
TESCHNER, George, State College of Virginia/Newport News VII-2
TYMIENIECKA, Anna-Teresa, World Phenomenology Institute, Massachusetts III-9
VEGA, Martin, Embassy of Nicaragua X-1
VICHOT, Roberto J., Texas A&M University VI-5
VILLANUEVA, Enrique, Instituto de Investigaciones Filosoficas, Mexico II-13
WAHL, Russell, Idaho State University II-12
WALSH, Joe, Stockton State College, New Jersey VIII-1
WALTERS, Gregory J., University of Ottawa, Canada III-6
WEINSTEIN, Mark L., Queens College, City University of New York IV-4
WEISS, Paul, Catholic University of America VII-4
WEISSMAN, David, City College, City University of New York III-1
WERHANE, Patricia, Loyola University of Chicago III-2
WESTPHAL, Merold, Fordham University IX-2
WESTRA, Laura, University of Toledo V-8
WHITE, Richard, University of San Diego II-1
WHITNEY, Gordon E., Princeton University IV-4
WILSON, Arnold, University of Cincinnati V-13
WILSON, Margaret, Princeton University VI-4
WINFIELD, Richard Dien, University of Georgia VII-2
WOLTER, Allan B., Old Mission, Santa Barbara, California X-4
WRIGHT, Kathleen, Haverford College, Pennsylvania II-3
YOUNG, Iris, Worcester Polytechnic Institute V-1

the review of

metaphysics

a philosophical quarterly

MARCH 1987 | VOL. XL, No. 3 | ISSUE No. 159 | $8.00

articles

ROBERT SOKOLOWSKI *Exorcising Concepts*
ARTHUR RIPSTEIN *Explanation and Empathy*
KENNETH G. FERGUSON *Truth Conditions for "Might" Counterfactuals*
ROBERT E. WOOD *Image, Structure and Content: On a Passage in Plato's Republic*
THOMAS J. BOLE, III *Contradiction in Hegel's Science of Logic*

critical studies

ROBERT B. PIPPIN *Blumenberg and the Modernity Problem*
JANICE FLOERSCH and Staff *Summaries and Comments*

philosophical abstracts *announcements*

Individual Subscriptions $21.00 Institutional Subscription $32.00 Student Subscriptions $10.00

Jude P. Dougherty, Editor
The Catholic University of America, Washington, D.C. 20064

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

New in Logic—1987/88

Ask us about the new tutorial software available with this edition of Hurley (Booth #35)!

Patrick J. Hurley
A CONCISE INTRODUCTION TO LOGIC, Third Edition
One of the most respected and popular logic texts in print today! An excellent blend of formal and informal topics. Now contains *hundreds of real-life examples and exercises.*
(Available January 1988)

Trudy Govier
A PRACTICAL STUDY OF ARGUMENT, Second Edition
A new edition to this successful approach to critical reasoning. Concentrates on good arguments as they occur in natural language.
(Available September 1987)

Trudy Govier
SELECTED ISSUES IN LOGIC AND COMMUNICATION
A thought-provoking anthology for critical thinking courses. Examines the connection between critical thought and the media in our society.
(Available September 1987)

Howard Kahane
LOGIC AND CONTEMPORARY RHETORIC: THE USE OF REASON IN EVERYDAY LIFE, Fifth Edition
The most widely used informal logic book keeps getting better! This new edition contains hundreds of NEW examples including such topics as "Iran-Gate," AIDS, and STARWARS.
(Available January 1988)

Zachary Seech
LOGIC IN EVERYDAY LIFE
A skills-oriented critical reasoning text that serves students well in the classroom and beyond. Friendly and direct with many exercises.
(Available September 1987)

New Software for Logic Students

LEARNING LOGIC: THE BASICS
Ron Messerich and Frank Williams
A helpful drill and practice tutorial for use with any critical reasoning text. *(Available January 1988)*

WADSWORTH

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIV, NO. 11, NOVEMBER 1987

PERSONAL AUTONOMY AND THE PARADOX OF FEMININE SOCIALIZATION*

PEOPLE are personally autonomous when their conduct is morally permissible and is not dictated by any technical rule, and when they are doing what they, as individuals, want to do. On this account, autonomous personal decisions might seem reducible to straightforward questions of maximizing desire satisfaction, but Freud and Marx have shown how opaque the divide between perceived desire and actual desire commonly is. Their insight exposes the inadequacy of understanding personal autonomy as nothing more than doing what one wants without undue interference from others, and it highlights the need for an account of the difference between doing what one wants and doing what one *really* wants. The autonomous self is not identical with the apparent self; it is an authentic or "true" self. Autonomous conduct expresses the true self.

Since modern social science numbers socialization among the prime threats to personal autonomy, a good way to get one's bearings in regard to this topic is to examine the effects of a socialization process that is widely believed antithetical to autonomy. Traditional feminine socialization—the set of practices which instills in girls the gentle virtues of femininity along with homespun feminine goals—is one such process.[1] According to many feminist scholars, feminine socialization is crucial to the persistence of women's subjection. Yet,

* To be presented in an APA symposium on Feminist Perspectives on Individual Choice, December 30, 1987. Kathryn Pyne Addelson will comment; see this JOURNAL, this issue, 628/9.

I would like to thank Kathryn Jackson for her helpful comments and the American Council of Learned Societies for the support of an ACLS/Ford Fellowship.

[1] I am confining my discussion to prevalent practices within the dominant cultures in Western industrialized nations.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

0022-362X/87/8411/0619$01.00 © 1987 The Journal of Philosophy, Inc.

Digitized by Arya Samaj Foundation Chennai and eGangotri

against the claim that feminine socialization suppresses personal autonomy, some feminist thinkers have celebrated what they regard as distinctively feminine styles of agency. I shall begin by focusing on the issues these diametric views of feminine socialization raise, and I propose to ask what sort of account of personal autonomy would be adequate to both of these views. Though I am taking feminist theory as my point of departure, I believe that the conception of personal autonomy implicit in this literature has general application, and that it is a more satisfactory one than we have heretofore had.

I. THE PROBLEM OF PERSONAL AUTONOMY IN FEMINIST THEORY

Much recent feminist theory has adopted the attractive, but somewhat paradoxical position that women have traditionally occupied oppressive, subordinate social positions, but that traditional women and their work should be respected as such. Feminists have not always subscribed to the latter thesis, however. In *The Second Sex,* Simone de Beauvoir[2] repeatedly declares that, as a result of their exclusion from public life and their socialization for domesticity, women in their present condition are "mutilated." Their work consists of deadly maintenance routines, and their creative potential is stunted. Whereas men are capable of transcendence, women are trapped in immanence—a state of passivity and objectification. Thus, de Beauvoir concentrates on explaining how women, who are capable of being men's equals, are systematically rendered inferior to men.

Though few feminists today would concur with de Beauvoir's virulent condemnation of women as mutilated beings, the basic line of argument which de Beauvoir developed remains in currency as an explanation of why women's status remains subordinate to men's despite the fact that women have had full political and economic rights for over two decades. It could be argued, of course, that prejudice against women lingers and that this era of legal equality is brief in comparison with the time needed to prepare for and establish careers. There is evidence, however, that women pursuing careers outside the home often suffer from conflicts regarding their sexual identity which never afflict men.[3] Discrimination cannot explain such ambivalence. To account for this phenomenon, some

[2] (New York: Bantam, 1961), pp. 274, 295, 306, 451, and 642.

[3] Matina Horner has argued that talented women whose prospects are excellent fear success and avoid success because they regard achievement as masculine and therefore associate success with loss of femininity. For helpful reviews of this research, see Lenore J. Weitzman, "Sex-role Socialization: A Focus on Women," in *Women: A Feminist Perspective,* Jo Freeman, ed. (Palo Alto, CA: Mayfield, 1984), pp. 202–204; and Janet Shibley Hyde, *Half the Human Experience: The Psychology of Women* (Lexington, Mass.: D.C. Heath, 1985), pp. 204–206.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

feminists have argued that differential childhood socialization continues to funnel girls into the psychology of dependency and altruistic devotion to others which is traditionally associated with femininity. If this is true, it is no surprise that women would resist acting aggressively and self-interestedly and would therefore be less successful in public vocations than their male peers. On this view, women have been crippled by a mind set instilled in them when they were children.

Still, the conclusion that women have been made inferior to men is by no means a palatable one. Not only does it strip countless homemakers of any dignity whatsoever, but also it fails to take account of the sensitivity and imagination that childcare requires. To redress this unbalanced view of women, many contemporary feminists have valorized the traditional feminine role. Hence, a second argument now in circulation points out that their domestic role has uniquely situated women to develop practices which focus on the particular needs of individuals and which maintain lasting relationships among people. Whereas men's capacities for calculated maneuvering and for achieving their goals are highly developed, women's capacities for intimate interaction and for resolving differences among people are highly developed. Since both types of capability are vital to social life, the ways in which women and men approach decision making must be regarded as equal, albeit different.

Although the argument deprecating women's autonomy is in tension with the argument celebrating it, they both ring true. Thus, it seems to me advisable to embrace both lines of argument. There is ample social-psychological evidence to the effect that women are less able to exert control over their lives than men, but the claim that feminine socialization altogether excludes most women from the class of autonomous agents is both morally repugnant and factually unsubstantiated. What is needed is an account of personal autonomy which comprehends the experiences of traditional women but which also acknowledges the liabilities that curtail these individuals' autonomy.

II. PERSONAL AUTONOMY AND THE TRADITIONAL WOMAN'S LIFE

Together, the two arguments I have sketched yield a composite portrait of traditional women as attentive to others' welfare, adept at orchestrating harmony within groups, and dependent on others for approval and direction. This is not particularly promising material on which to base an account of personal autonomy. Since autonomous people are self-governing, the self must be a pivotal element in any conception of personal autonomy. But, in this portrait of the traditional woman, the self is conspicuously absent. Furthermore, no

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

account of personal autonomy could be complete without an explanation of how autonomous individuals govern their lives. But this portrait of the traditional woman depicts an individual more deferential to others than assertive of her own concerns.

To salvage whatever personal autonomy the traditional woman may have, an account of personal autonomy must accommodate three features of the traditional woman's position: (1) strongly directive prior socialization; (2) deep emotional ties to other people; (3) a home-centered, rather than a work-centered, orientation. Now, the initial response to this list might be to object that these are the very conditions that prevent traditional women from attaining personal autonomy. To forestall this objection, I shall sketch some reasons for countenancing these constraints on personal autonomy. My contention will be that our normal attribution of some degree of personal autonomy to most men obliges us to allow that these features of the traditional woman's life do not necessarily rule out personal autonomy.

Taking the third point first, it is puzzling why, from the standpoint of personal autonomy, life in the public sphere should be privileged over life in the private sphere. To be autonomous, people must conduct their lives in accordance with the dispositions of their authentic selves. Since individuals differ, it is unlikely that one sort of environment would give every self maximal opportunity to flourish. Indeed, the public sphere is itself a highly variegated setting that compasses diverse occupations befitting different individuals. Thus, it seems appropriate to view the home as yet another arena in which the true self might find expression.

Still, it might be countered that, since homemaking is unsalaried labor, the homemaker is economically dependent on her husband, and this dependency prevents the traditional woman from acting on her views when this would antagonize her husband. However desirable economic self-sufficiency may be from the standpoint of personal autonomy, all employees are dependent on their bosses. Since they are sometimes forced to choose between their convictions and their jobs, working for pay does not guarantee independence. Moreover, some benefactors intrude so little in their dependents' lives that these beneficiaries' autonomy is in no way compromised. Although the homemaker's reliance on her husband's income could undermine her autonomy, she may be no more constrained—indeed, she may be less constrained—than many paid employees. Since economic dependency usually leaves considerable latitude for personal choice, requiring freedom from economic encumbrances as a precondition of personal autonomy would be too restrictive. Still, in

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

assessing any given individual's autonomy, it is important to ascertain the impact of economic necessity on that person's choices.

At this point, it might be urged that the second of the conditions I mentioned—that is, deep emotional ties to other people—compounds the traditional woman's dependency on her husband in a way that beleaguers few wage-earning employees. Not only does economic self-interest militate against the traditional woman's opposing her husband, but also her love for him keeps her subservient.

Although it may be an added disadvantage to love the person upon whom one's livelihood depends, it is a mistake to think that the commitments involved in profound affection are peculiar to the traditional woman's attachments. Excluding love from the autonomous life would imply the exclusion of a whole range of life plans which an autonomous person could reasonably take up. Success in many of the more stimulating, prestigious, and well-paid careers involves identification with a set of standards and objectives ordained by the profession, not by each of its individual practitioners. Similarly, political advocacy or dedication to other sorts of causes is often passionate and often involves relinquishing some of one's prerogatives in the service of one's convictions. Certainly, people can become so absorbed in their careers or in social movements that they cease to be capable of discerning a mismatch between their activities and their selves. But, since it is hard to believe that shallow commitments are the only ones which can give expression to anyone's authentic self, it is necessary to discover how an equilibrium can be maintained between the integrity of the self and the people and projects the self embraces.

Returning to the problem presented by the traditional woman, it is a truism that her loving attachment to her family can supplant her self. Yet, it is not clear that her affection altogether eclipses her self, and it would be a severely impoverished conception of the authentic self which denied that love could be an expression of it. On the assumption that few people are so unfortunate as to love no one and to be unloved, either personal autonomy is reserved for those whose lives are emotionally empty, or personal autonomy is possible in the context of ongoing love. Although the commitments and interdependencies implicit in love complicate the problem of autonomy, it seems better to contend with these complexities than to trivialize autonomy.

Finally, if anything can, strongly directive socialization seems to controvert personal autonomy. Since the traditional woman has been assiduously groomed for the feminine role from the moment she was first wrapped in a pink blanket, her fulfilling her duties as

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

housewife and mother seems a paradigmatic case of someone's doing what others want and expect one to do, regardless of one's own desires. In contrast, men seem blessed with options: they may choose from an array of careers, and bachelorhood does not carry the connotations of rejection and loneliness which stigmatize spinsterhood. Nevertheless, masculine socialization is powerfully directive. Little boys are taunted for "sissyish" behavior, and, through identification with their fathers or other male models, they are firmly guided into the role of paterfamilias. For men as well as women, then, personal autonomy, if it is ever attained, must be attained within a context of choice tempered by childhood socialization.

I do not mean to suggest that the traditional woman is somehow an ideal exemplar of personal autonomy. Quite the contrary, her autonomy is seriously threatened in a number of ways. I do want to insist, however, that the standard doubts about her credentials as an autonomous agent can be cogently lodged against most men, and that these doubts do not warrant the conclusion that the traditional woman altogether lacks autonomy.

III. THREE FORMS OF AUTONOMY

To leave room for personal autonomy despite pervasive childhood role preparation coupled with compelling incentives to accede to social norms, it must be possible for people to act autonomously in isolated situations, and to adopt some projects and policies autonomously without having control over the basic direction of their lives. In other words, it must be possible for a life to contain pockets of autonomy and threads of autonomy which do not add up to an autonomous life. In defense of this conception, I shall argue that viewing personal autonomy as an all-or-nothing phenomenon is misguided in several respects. Specifically, I shall urge that the scope of programmatic autonomy compasses narrow as well as global issues, that episodic autonomy is intelligible without programmatic autonomy, and that a measure of personal autonomy can be gained through partial insight into one's authentic self.

Programmatically autonomous people pose and answer the question: "How do I want to live my life?" But this question cannot be addressed unless it is broken down into subsidiary questions, such as: "What line of work do I want to get into?" "Do I want to have children?" "Do I care more about material gain or spiritual values?" and the like. A person is programmatically autonomous when that person is carrying out a life plan that embodies that person's answers to this family of questions.

Still, it is evident that questions about the overall course of one's life can be pegged at different levels of generality. With respect to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

procreation, for example, one can ask questions as broad as whether one wants to have any children at all, or as narrow as which contraceptive method to use. Moreover, autonomy at the highest level of generality is not necessary for autonomy at lower levels. Indeed, since less is at stake, people may find it easier to discern what they really want in regard to narrower matters. Although people who query themselves regarding the most general aspects of their life plans are plainly more autonomous than people who by and large conform to social expectations and only attend to themselves regarding the minutiae of their lives, the latter may display narrowly programmatic autonomy.

A person is episodically autonomous when, in a particular situaton, that person asks "What do I really want to do now?" and acts on the answer. Here, the individual is not formulating a long-term plan or setting a policy to be applied in other circumstances; the question is confined to a single action. Autonomy is enhanced to the extent that the convictions and attitudes entering into the decision have been previously examined and endorsed. Yet, people can gain a measure of autonomy by addressing situation-specific questions that occasion introspective reflection. For they get to know themselves better, and they give greater expression to their own beliefs and desires than someone who mindlessly apes convention or caves in to others' wishes.

The traditional woman's personal autonomy is suspect, because she has been inducted into the feminine role willy-nilly. But traditional women cannot always rely upon popular formulas or feminine norms to handle the predicaments they confront. Thus, a traditional woman who summons the courage to demand that her child's teacher show respect for values that she cherishes may be acting in an episodically autonomous fashion. This is not to imply that individuals who never ask themselves whether the general life plans they are pursuing are the ones they really want to pursue are nonetheless fully autonomous, but it is to affirm that such people may have more personal autonomy than one might suppose if one never examined the details of their lives.

We have seen that the autonomy of a life is a matter of degree inasmuch as distinct undertakings can be autonomous, though the overarching life plans that subsume them are not autonomous. It is equally important to recognize that the autonomy of any particular decision is also a matter of degree. Consider the case mentioned above of the mother protesting a teacher's contempt for the mother's values in her child's school. To be fully autonomous, the mother's action must not only be the most sincere and effective

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

expression of her concern, but also it must articulate values that she has independently accepted. Yet, someone could fall short of this ideal and still evidence some degree of personal autonomy.

Suppose that the mother is a Christian fundamentalist who is angered by the teacher's glib dismissal of creationism, and suppose further that she is a fundamentalist preacher's poorly educated daughter who has never been exposed to less benighted theological doctrines and who has never questioned her faith. Plainly, her conduct is heteronomous to the extent that it voices beliefs that she regards as immune to criticism. Yet, insofar as she succeeds in conveying the outrage that she genuinely feels in a manner that enables the teacher to fathom and to respect the intensity of her conviction, she is exercising control over her life. In this limited manner, she is acting autonomously. Thus, partial access to and expression of the self must not be overlooked as a form that autonomy can take.

Autonomy can be compromised in a number of ways. Although lack of critical rationality in considering one's traits, convictions, and plans subverts personal autonomy, lack of sensitivity to one's feelings, lack of inventiveness in conceiving plans and actions, and lack of determination to carry out one's decisions undermine personal autonomy, too. Accordingly, personal autonomy must be gauged along a number of dimensions, and fragmentary autonomy can sometimes be descried in lives that are, in many important respects, heteronomous.

IV. A COMPETENCY APPROACH TO PERSONAL AUTONOMY

If we accept that there are at least three lesser forms of autonomy, it is necessary to ask what conception of autonomy would account for them. Various theories of autonomy come to mind. One possibility would be that, when a person achieves episodic, narrowly programmatic, or partial access autonomy, she in some measure defies her prior socialization. A second possibility would be that this person occasionally gives some degree of expression to those of her qualities which are innate. A third possibility would be that this individual sometimes decides more or less definitely to identify with her effective desires. None of these proposals is devoid of appeal. I would urge, however, that, whatever account of the autonomous self one adopts, it will be necessary to introduce a repertory of skills to explain how this self finds an outlet.

Since one must exercise control over one's life to be autonomous, autonomy is something that a person accomplishes, not something that happens to a person. If insanity enabled one to overcome one's socialization, to project one's innate qualities, or to identify with one's desires, one would not thereby become autonomous. Only if

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

one expresses one's authentic self by exercising assorted introspective, imaginative, reasoning, and volitional skills would we agree that the resulting conduct was autonomous.

Consider how people might achieve programmatic autonomy in the broadest sense—that is, how people could autonomously direct the overall course of their lives. Roughly, this sort of autonomy requires that people vividly envisage different life plans and seriously entertain them. Attuned to the feelings evinced by their alternatives, they must interpret these feelings correctly, and evaluate them critically. In light of relevant factual information, they must assess the practicality of these options. Likewise, they must judge the merits of sundry proposals in terms of their other values. Moreover, autonomous people must recognize and act on signs of discontent with previous decisions. They must be prepared to acknowledge inner change and must be willing to modify their plans in response to such change. Finally, comprehensive programmatic autonomy requires that people be ready to resist the unwarranted demands of other individuals along with conformist societal pressures, and that they be resolved to carry out their own plans. In sum, to the extent that individuals survey their options guided by their self-scrutinized feelings, values, goals, and the like, and then marshall the determination to follow their own counsel, they live autonomously.

Personal autonomy, then, can only be achieved through the exercise of a repertory of coordinated skills, which constitutes autonomy competency. People are prevented from extending their autonomy to global programmatic decisions when they lack proficiency in one or more of these skills or when these skills are ill-coordinated. These deficiencies can explain the traditional woman's limitation to sporadic and gradated forms of autonomy—narrowly programmatic autonomy, episodic autonomy, and partial access autonomy. Not only does traditional feminine socialization propel women into the low status role of housewife and mother, but also it curtails the development of autonomy competency.[4] Thus, the efficacy of this early childhood experience is redoubled.

This view of autonomy and feminine socialization enables us to see how both of the feminist positions with which I began can be correct. The traditional woman achieves a measure of personal autonomy, for she does not altogether lack autonomy competency. But, because the traditional woman's autonomy competency is underdeveloped, she cannot resolve the conflict between her internalized image of

[4] See part 3 of my *Self, Society, and Personal Choice* (New York: Columbia, forthcoming).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

womanliness and her career aspirations. In addition, this view of autonomy and feminine socialization explains why feminists are justifiably suspicious of the professed fulfillment of many traditional women. Since traditional women do not use autonomy skills adeptly, there is no reason to believe that they are doing what they really want to do. Conversely, however, if an adult who has been raised to assume the tasks of housekeeping and parenting embraces this role, feminists would have no grounds for complaint provided that the individual is skilled in autonomy competency. Since, in principle, the traditional feminine role could be the object of autonomous choice, feminists cannot presume to exclude it as a candidate life plan. Yet, insofar as feminine socialization impedes the development and the exercise of autonomy competency, feminists must seek to overhaul these socialization practices.

DIANA T. MEYERS

University of Connecticut/Storrs

AUTONOMY AND RESPECT*

The classic feminist arguments in the United States have been rooted in the liberal ideal of personal autonomy. Diana Meyers finds a paradox here, for autonomy has been male defined in terms of public sphere activities in a way that denigrates the domestic sphere and shows no respect for women's traditional abilities. In her attempt to resolve the paradox, she offers an ambitious and useful analysis by taking autonomy to be a way of living in harmony with one's true self, displayed as the exercise of competences in making life plans, making immediate choices, and having access to oneself. The analysis integrates recent feminist research in several fields, and it brings mainstream philosophical discussions of autonomy to a new sophistication.

The feminist paradox is set in the contrast between autonomy and heteronomy, which Meyers interprets in terms of socialization in a way that both sets the problem and offers the clue to her solution. Socialization for women's traditional role may impede the full development and the exercise of autonomy competences—and so feminists must seek to overhaul those socialization practices. Meyers would presumably say the same thing about socialization to men's

* Abstract of a paper to be presented in an APA symposium on Feminist Perspectives on Individual Choice, December 30, 1987, commenting on a paper by Diana T. Meyers, this JOURNAL, this issue, 619–628.

0022-362X/87/8411/0628$00.50 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

traditional role. Her analysis is detailed enough so that we can see what skills need to be installed—introspective, imaginative, reasoning, and volitional skills.

Meyers's own analysis is liberal, and it offers a more sophisticated basis for mainstream feminism. I fear it carries contemporary liberal biases—which accounts for its usefulness and practicality in our liberal society. The idea of autonomous individuals suits the way we elect officials, judge felons, hire workers, pay wages and taxes, register births and marriages, and enroll and grade students. The liberal institutions today give our authority and pride of place not to males, but to professionals. Certainly, much of the socialization takes place in schools, with trained teachers (including philosophers), career counselors and therapists, sex educators (for family planning), and appropriate textbooks. Parents need to buy toys and games that encourage autonomy skills, and they themselves need books and workshops on how to raise autonomous children. I do not say that Meyers insists on this, but this is the paradigm, and it is the one under attack in current skirmishes between liberal humanists and fundamentalists.

The problem is that Meyers has worked hard to show respect for the traditional abilities and traditional lives of women. But liberal analyses do not respect the worlds and lives of many men, women, and children, and they leave many of our own secular, white, professional biases unexamined. Feminists are aware of race and class biases—they worry about charges that the women's movement is white and "middle class." But there are also liberal, secular biases. The fundamentalist mother is faulted for accepting creationism uncritically, but we do not ask about the schoolteachers' acceptance of Darwin—evolutionary theory is true, is it not? In one form or another, the spector of elitism re-emerges.

The notion of socialization as Meyers uses it and the liberal separation of individual and society are prime sources of these problems, and they are incorporated in the sociology and psychology that philosophers tend to rely upon. In my comment, I shall briefly give an alternative account of people in society, drawn from interactionist sociology. This account shows the divisions between public and private and the very separation of individual and society to be folk concepts that operate to maintain social hierarchies and elites. I use the interactionist sociology as the basis of a moral notion of respect which lets us consider Meyers's paradox and see her analysis of autonomy in a new light.

KATHRYN PYNE ADDELSON

Smith College

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

MIND MATTERS*

> Who knows what I want to do? . . . Isn't it all a question of brain chemistry, signals going back and forth, electrical energy in the cortex? . . . Some minor little activity takes place somewhere in this unimportant place in one of the brain hemispheres and suddenly I want to go to Montana or I don't want to go to Montana. Maybe it's just an accidental flash in the medulla and suddenly there I am in Montana and I find out I really didn't want to go there in the first place . . . It's all this activity in the brain and you don't know what's you as a person and what's the brain and what's some neuron that just happens to fire or just happens to misfire . . .—Don DeLillo, *White Noise*

CONSIDER the following, admittedly imprecise claims:

1) The mental and the physical are distinct.
2) The mental and the physical causally interact.
3) The physical is causally closed.

Much can be said in favor of each of these. In support of (1), we can point to the failure of attempts to reduce the phenomenal and the intentional to the physical, and to arguments from Descartes to Donald Davidson which purport to show that such reductions are, in principle, impossible. (2) is supported by our everyday experience and by various theories of perception and action. (3) means that every physical event or fact has, in its causal history, only physical events and facts. Both (3) and its cousin:

3′) All causation is reducible to, or grounded in, physical causation,

where 'grounded' means, roughly, that causal relations supervene on noncausal physical facts and laws, have seemed to many philosophers to be supported by the development of the sciences.

The trouble is that it seems (1), (2), and (3) are incompatible. To be a bit more definite, consider their application to events. (1) then says that no mental event is a physical event; (2), that some mental events cause physical events and vice versa; and (3), that all the causes of physical events are physical events. The inconsistency is obvious. If

* To be presented in an APA symposium of the same title, December 30, 1987. Jerry Fodor will comment; see this JOURNAL, this issue, 642.

Thanks are due to Jonathan Adler, John Biro, Paul Boghossian, Donald Davidson, Fred Dretske, Ray Elugardo, Jerry Fodor, Richard Foley, Terry Horgan, Brian McLaughlin, Alexander Rosenberg, Stephen Schiffer, and John Searle.

0022-362X/87/8411/0630$01.30 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

mental events are distinct from physical events and sometimes cause them, then obviously the physical is not causally closed. The dilemma posed by the plausibility of each of these claims and by their apparent incompatibility is, of course, the mind–body problem.[1]

Our primary concern here is how Davidson's[2] account of the relation between the mental and the physical, which he calls "anomalous monism" (*AM*), attempts to resolve the dilemma. AM consists of the following three theses:

4) There are no strict psychophysical or psychological laws and in fact all strict laws are expressed in a purely physical vocabulary (the anomalousness of the mental).
5) Mental events causally interact with physical events.
6) Event *c* causes event *e* only if there is a strict causal law which subsumes *c* and *e* (entails that *c* causes *e*) (the nomological character of causality).

(4) is a version of (1). It is commonly held that a property expressed by *M* is reducible to a property expressed by *P* (where *M* and *P* are not analytically connected) only if there is an exceptionless bridge law that links them.[3] So it follows from (4) that (intentional) mental and physical properties are distinct.[4] (6) says that *c* causes *e* only if there are singular descriptions *D* of *c* and *D′* of *e* and a strict causal law *L* such that *L* and "*D* occurred" entail "*D* caused *D′*" ("Causal Relations," p. 158). (6) and the second part of (4) entail that physical events have only physical causes and that all event causation is physically grounded.[5]

The notion of a law being *strict* figures prominently both in Davidson's affirmation of the distinctness of the mental and the physical and in his account of causation. Davidson's notion of a strict law is

[1] Similar characterizations of the mind–body problem can be found in J. L. Mackie, "Mind, Brain and Causation," *Midwest Studies in Philosophy*, VI (1979): 19–29; and Anthony Skillen, "Mind and Matter: A Problem that Refuses Dissolution," *Mind*, XCIII (1984): 514–526.

[2] This view is given in three places in Davidson's *Essays on Actions and Events* (New York: Oxford, 1980); at the beginning and end of "Mental Events," pp. 208, 223; and in "Psychology as Philosophy," p. 231. Where nothing else is said, all page references in the text of our paper are to this book.

[3] Davidson's argument against psychophysical laws is restricted to laws whose psychological predicates express propositional attitudes.

[4] We shall typically speak of features, aspects, and properties of events. For present purposes, however, unless we indicate otherwise, what we say can be recast in terms of events satisfying descriptions or predicates.

[5] Davidson never provides an example of a strict causal law. And there are some philosophers who think his account of causation is much too stringent, because there may be too few strict causal laws. (The best candidates for such laws are basic laws of quantum mechanics.) It is not our aim here to defend Davidson's metaphysical account of causation.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

best explained by contrast with nonstrict laws. A nonstrict law is a generalization that contains a *ceteris paribus* qualifier that specifies that the law holds under "normal or ideal conditions," where the relevant notions of normal or ideal are specified by the theoretical context of the law. The generalizations one finds in the special sciences are mostly of this kind. In contrast, a strict law is one that contains no *ceteris paribus* qualifiers; it is exceptionless not just *de facto* but as a matter of law. A nonstrict law may be improved upon by explicitly including some of its *ceteris paribus* conditions in its antecedents. Davidson's view is that psychophysical laws of the form —whenever a person is in physical state *P*, then he is in intentional state *M*—are *essentially* nonstrict. That is, no matter how many conditions are added to the antecedent, short of trivializing the generalization, it will not be strict.[6]

Given the parallel between (4)–(6) and (1)–(3), it may seem that the former are also incompatible. But they are not. Davidson shows that they all can be true if (and only if) mental events are identical to physical events ("Mental Events," p. 215). Let us say that an event *e* is a physical event just in case *e* satisfies a basic physical predicate (that is, a physical predicate appearing in a strict law). Since only physical predicates (or predicates expressing properties reducible to basic physical properties) appear in strict laws, it follows that every event that enters into causal relations satisfies a basic physical predicate. So, those mental events which enter into causal relations are also physical events.

AM is committed only to a partial endorsement of (1). The mental and physical are distinct insofar as they are not linked by strict law —mental properties are not reducible to physical properties—but they are not distinct insofar as mental events are physical events. This being so, one might wonder whether AM also only partially endorses claims (2) and (3). In fact, Davidson's views have been criticized precisely on the point of (2). Ernest Sosa[7] writes: "I conclude that . . . anomalous monism is [not] really compatible with the full content of our deep and firm conviction that the mind and body each acts causally on the other" (*ibid.*, p. 278). Ted Honderich[8] goes even

[6] For an explication and defense of Davidson's arguments for the impossibility of strict psychophysical laws, see Jaegwon Kim, "Psychophysical Laws," *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, LePore and Brian McLaughlin, eds. (London: Basil Blackwell, 1986), pp. 369–386, and also the introduction by McLaughlin in the same volume. Cf. also LePore and Loewer, "Davidson and the Anomalousness of the Mental," *Philosophical Perspectives on the Philosophy of Mind*, J. Tomberlin, ed., forthcoming.

[7] "Mind-body Interaction and Supervenient Causation," *Midwest Studies in Philosophy*, IX (1984): 271–282.

[8] "Smith and the Champion of the Mauve," *Analysis*, XLIV, 2 (1984): 86–89.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

further charging that AM is really a form of epiphenomenalism: "I went on . . . to claim that [AM] was epiphenomenalist; it did not make the mental as mental an ineliminable part of the explanation of actions" (*ibid.*, p. 88).

If Honderich means that Davidson's views are committed to epiphenomenalism with respect to mental events, he is clearly mistaken, since, according to AM, mental events do cause other events. They *are* physical events and so can, like any event, have consequences. It is rather that, on AM, as he puts it, the mental *as* mental—some writers use the expressions "*qua* mental" and "*in virtue of being* mental"—is causally irrelevant. In defense of Davidson, one might reply that, although it is correct to say it is not *c* as mental that causes *e*, this has nothing to do with any epiphenomenalism on the part of the mental, but simply reflects the fact that it is not events *as* mental or *as* physical or *as* anything else which cause other events. Causation is a relation between events, not between events *as F*s. It seems to Davidson's critics, however, to make sense to distinguish some features of an event as causally relevant and others as causally irrelevant. It is this distinction which underlies the locution that it is *c as F* (not *as F′*) that causes *e* (to be *G*). Sosa and Fred Dretske[9] illustrate their understanding of the distinction in the following passages, respectively:

> A gun goes off, a shot is fired and it kills someone. The loud noise is the shot. Thus if the victim is killed by the shot it is the loud noise that kills the victim. . . . In a certain sense the victim is killed by the loud noise. Not by the loud noise as a loud noise but only by the loud noise as a shot, or the like . . . . The loudness of the shot has no causal relevance to the death of the victim. Had the gun been equipped with a silencer the shot would have killed the victim just the same (*op. cit.*, pp. 277/8).

> Meaningful sounds, if they occur at the right pitch and amplitude, can shatter glass, but the fact that these sounds have a meaning is irrelevant to their having this effect. The glass would shatter if the sounds meant something completely different or if they meant nothing at all (*op. cit.*).

[9] "Reasons and Causes," manuscript presented at the Chapel Hill Colloquium. For similar characterizations and examples of causal relevance, see Ted Hoderich, "The Argument for Anomalous Monism," *Analysis*, XLII, 1 (January 1982): 61; John Searle, *Intentionality* (New York: Cambridge, 1983), pp. 155–157; Elizabeth Anscombe, "Causality and Extensionality," *Causation and Conditionals*, E. Sosa, ed. (New York: Oxford, 1975), p. 178; and Peter Achenstein, "The Causal Relation," *Midwest Studies in Philosophy*, IV (1977): 868.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Sosa, Honderich, Jaegwon Kim, Dretske, (among others)[10] think that, once we have made the distinction between the causally relevant and irrelevant features of an event, we will see that it is a consequence of AM that mental features are never causally relevant. Why is the causal irrelevance of the mental supposed to be entailed by AM? Kim[11] reasons as follows:

> Consider Davidson's account: whether or not a given event had a mental description . . . seems entirely irrelevant to what causal relations it enters into. Its causal powers are wholly determined by the physical description or characteristic that holds for it; for it is under its physical description that it may be subsumed under a causal law. And Davidson explicitly denies any possibility of a nomological connection between an event's mental description and its physical description that could bring the mental into the causal picture (*ibid.*, p. 267).

The argument is that, since, according to AM, *c* causes *e* only if there is a strict law that subsumes *c* and *e* and since strict laws contain only physical (never mental) predicates, it follows that the mental features of events *c* and *e* are irrelevant to whether they are causally connected. The physical features of events suffice to fix, given the strict laws, *all* causal connections. Mental features neither suffice nor are required to fix causal connections. The argument is powerful. The conclusion the authors draw from it is that on AM *the mind does not matter;* that a neural event has a certain intentional content is as irrelevant to its effect as the fact that the sounds are meaningful is to the sounds causing the glass to break.

But is this criticism of AM correct? We claim that it is not, and that it rests on a simple, but perhaps not obvious, confusion. The confusion is between two ways in which properties of an event *c* may be said to be causally relevant and irrelevant. Consider the following locutions:

a) Properties *F* and *G* are relevant$_1$ to making it the case that *c* causes *e*, and

b) *c*'s possessing property *F* is causally relevant$_2$ to *e*'s possessing property *G*.

[10] Others who have argued that AM is epiphenomenalist include: F. Stoutland, "The Causation of Behavior," *Essays on Wittgenstein in Honor of G. H. von Wright* (Acta Philosophica Fennica, 1976), p. 307; Dagfinn Føllesdal, "Causation and Explanation: A Problem in Davidson's View on Action and Mind," *Action and Events: Perspectives on the Philosophy of Donald Davidson, op. cit.*, p. 315; Mark Johnston, "Why the Mind Matters," *ibid.*, p. 423; and Skillen, *op. cit.*, p. 520.

[11] "Epiphenomenal and Supervenient Causation," *Midwest Studies in Philosophy*, IX (1984), 257–270.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

We will say that (a) holds iff c has F and e has G, and there is a strict law that entails Fs cause Gs. It is in this sense that it is c's having F and e's having G "make it the case" that c causes e. Relevance$_2$ is a relation among c, one of its properties F, e, and one of its properties G. It holds when c's being F brings it about that e is G. We shall argue that those who charge AM with epiphenomenalism are guilty of confusing relevance$_1$ with relevance$_2$.

None of the authors we have been considering defines the sense of causal relevance they have in mind when they accuse AM of rendering the mental causally inefficacious. Their discussions, though, do suggest a *test* for causal irrelevance. Recall Sosa's remark that "had the gun been equipped with a silencer it would have killed the victim just the same" (278); and Dretske's remark that "the glass would shatter if the sounds meant something completely different." So it may be that Sosa and Dretske (and others) think that AM entails the causal irrelevance of the mental, because they think that it entails the falsity of such mentalistic counterfactuals as: if Fred had not believed that Jerry would attend the conference, he would not have come.

In view of this counterfactual test for causal irrelevance$_2$, we suggest that the authors who propose it may have in mind the following characterization of causal relevance$_2$.[12]

(I) c's being F is causally relevant$_2$ to e's being G iff
  i. c causes e.
  ii. Fc and Ge.
  iii. $-Fc > -Ge$.
  iv. Fc and Ge are logically and metaphysically independent.[13]

Condition (iv) is intended to exclude cases in which the connection between F and G is conceptual/metaphysical rather than causal, e.g., c's being the cause of e is causally relevant$_2$ to e's being caused by c, when c does cause e.

The heart of our response to the claim that AM is committed to epiphenomenalism is this: AM entails that mental features are caus-

[12] While many philosophers appeal to the notion of causal relevance, it is far from clear that there is a single or well-characterizable notion that underlies the locution that c qua F causes e to be G. We are here interested only in sketching enough of an account to refute the charge that AM is committed to epiphenomenalism. Anyone interested in a thorough explication of causal relevance would have to show how to accommodate familiar difficulties involving pre-emption, overdetermination, and so on. But these are problems which confront every account of causation and we will not discuss them here.

[13] c's being F and e's being G are metaphysically independent, iff there is a possible world in which c (or a counterpart of c) is F but e (or a counterpart of e) fails to occur or fails to be G and vice versa.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ally irrelevant$_1$, but does not entail that they are causally irrelevant$_2$. Before arguing these claims, we need to discuss the interpretation of the counterfactual:

(*Q*) If event *c* were not *F*, then event *e* would not be *G*.

We will adopt the Lewis–Stalnaker[14] account of counterfactuals, according to which $A > B$ is true iff B is true at all the worlds most similar to the actual world at which A is true (or A is true at no such world). We will suppose that an event *e* that occurs at the actual world may occur or have counterparts that occur at others. '*c*' and '*e*' are to be understood as rigid designators of events. In evaluating (*Q*), we need to look at the most similar worlds to the actual world at which *c* fails to be *F*. *c* may fail to be *F* at *w* either by existing there and not being *F* or failing to occur at *w* (or have a counterpart) at all. (*Q*) is true just in case the most similar worlds at which counterparts to *c* fail to have *F* or at which *c* fails to have a counterpart are such that counterparts to *e* fail to have *G* or *e* fails to have a counterpart.

The irrelevance$_1$ of the mental follows immediately from the definition of relevance$_1$ and from AM's (4) and (6). The irrelevance$_1$ of psychological predicates, however, is perfectly compatible with the truth of counterfactuals $-Fc > -Ge$, where *F* and *G* are predicates that do not occur in strict laws. That is, the set of strict laws and basic physical facts do not by themselves settle the truth values of counterfactuals.

We can see that this is so as follows: consider the set of worlds *W* at which all the strict laws hold. (This set includes the actual world α.) Until a *similarity order*, $\geq\alpha$, is placed on *W*, the truth values of almost all counterfactuals are indeterminate. Only those counterfactuals $A > B$ such that the strict laws and noncounterfactual statements true at α entail $A \rightarrow B$ or $-(A \rightarrow B)$ have determinate truth values, since any similarity ordering $\geq\alpha$ will make the former true and the latter false. This is just the lesson of Nelson Goodman's[15] failed attempts to analyze counterfactuals in terms of laws. What Goodman found is that laws and noncounterfactual truths are themselves not sufficient to settle the truth value of any but a limited set of counterfactuals. It follows that the truth of counterfactuals of the sort needed to establish causal relevance$_2$ (since neither they nor their negations are entailed by the strict physical laws and noncounterfactual truths) are compatible with AM.

[14] See David Lewis, *Counterfactuals* (Oxford: Blackwell, 1973); and Robert Stalnaker, "A Theory of Conditionals," *Studies in Logical Theory*, N. Rescher, ed. (New York: Oxford, 1968). There are differences between the two accounts irrelevant to our discussion.

[15] *Fact, Fiction and Forecast*, 4th ed. (Cambridge, Mass.: Harvard, 1982).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Of course, it is one thing to show that mentalistic counterfactuals are compatible with AM. It is quite another thing to produce an account of what makes these counterfactuals true and also show that this account is compatible with AM. The question of what makes counterfactuals true is a general one which concerns all counterfactuals and not just mentalistic ones. We shall briefly address it toward the end of our discussion.

To this point, we have shown that, if (I) supplies sufficient conditions for causal relevance$_2$, then there is no incompatibility between AM and the causal relevance$_2$ of the mental.[16] This is important, since, as we have seen, many of Davidson's critics seem to think there is such an incompatibility. There are two further related questions we need to address. One is whether causal irrelevance$_1$ alone is sufficient to sustain a charge of epiphenomenalism. A second is whether there are some further conditions on (I) such that, once added, AM does entail the causal irrelevance$_2$ of the mental.

Why would anyone think that irrelevance$_1$ of the mental entails epiphenomenalism? Honderich[17] formulates a principle he calls "the principle of the nomological character of causally relevant properties," according to which *c*'s having *F* is causally relevant to *e*'s having *G*, iff there is a law of the form *F*s cause *G*s (62). If one thinks, as Honderich does, that AM implies that psychological predicates never appear in causal laws, then one might conclude that psychological features have no causal role to play and indeed that psychology could not be a science. But, as Davidson has been careful to observe (240), there may very well be psychological and psychophysical causal laws that support counterfactuals and other subjunctive conditionals; it is just that such laws cannot be *strict*. If Honderich intends for the principle of nomological relevance to include nonstrict as well as strict laws, then AM is compatible with the causal relevance (in Honderich's sense) of psychological properties. If he intends for the principle to include only strict laws, then it is an unacceptable principle. It is implausible that there are any strict laws linking 'is a match striking' with 'is a match lighting'. So, on the strict law construal of

[16] Although there is a tradition in the philosophy of action arguing that there are conceptual connections between propositional attitudes and actions, this does not entail that particular propositional attitude properties are conceptually connected. For example, suppose that John believes that Mary is across the street and, for this reason, waves his hand. Let *c* be John's thought, *e* his action, *F* the property of his believing Mary is across the street, and *G* the property of being a waving hand. Clearly, we can have *c*'s being *F* causally relevant$_2$ to *e*'s being *G*, since *c*'s being *F* can obtain without *e*'s being *G* and vice versa in some metaphysically possible world.

[17] "The Argument for Anomalous Monism," *Analysis*, XLII, 1 (January 1982): 59–64.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Honderich's principle, being a match striking is not causally relevant to the match's lighting. On this construal, Honderich's principle would render virtually all properties of events causally irrelevant$_2$. This certainly seems wrong.

In arguing that AM entails the causal irrelevance of the mental, some authors have suggested a strengthened account of causal relevance$_2$. For example, Sosa writes:

> I extend my hand because of a certain neurological event. That event is my sudden desire to quench my thirst. Thus, if my grasping is caused by that neurological event, it's my sudden desire that causes my grasping . . . Assuming the anomalism of the mental, though extending my hand is, in a certain sense, caused by my sudden desire to quench my thirst, it is not caused by my desire qua desire but only by desire qua *neurological* of a certain sort. . . . [T]he being a desire of my desire has no causal relevance to my extending my hand (if the mental is indeed anomalous): *if the event that is in fact my desire had not been my desire but had remained a neurological event of a certain sort, then it would have caused my extending my hand just the same* (277/8, our emphasis).

This passage suggests the following as a sufficient condition for causal irrelevance$_2$:

(II) c's being F is causally irrelevant$_2$ to e's being G, if there is a property $F*$ of c such that $(F*c \ \& \ -Fc) > Ge$ holds nonvacuously.

Even when $-Fc > -Ge$ holds, there may be a property $F*$ of c such that $(F*c \ \& \ -Fc) > Ge$. In this case, it may seem that it is in virtue of c's being $F*$, not F, that e is G. When this holds, we will say that $F*c$ "screens off" Fc from Ge. Converting (II) into a necessary condition for causal relevance$_2$ and adding it to (I), we obtain the following proposal:

(III) c's being F is relevant$_2$ to e's being G iff the conditions in (I) are satisfied and there is *no* property $F*$ of c such that $(F*c \ \& \ -Fc) > Ge$ holds nonvacuously.

Sosa seems to think that it follows from AM that c's being a certain neural state, Nc, *screens off* c's being a desire to quench thirst, Mc, from e's being an extending of the hand, Be. More generally, he seems to think that neural properties screen off intentional mental properties. Presumably, Sosa thinks that this follows from AM, because he thinks there are strict laws connecting neural properties with behavioral properties. Since mental properties are not reducible to neural properties, it follows that there are physically possible worlds in which Nc, Mc, and in *all* such worlds Be.

It is not at all clear that there are strict laws connecting neural

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

properties with mental properties (and so that AM *entails* that the neural property screens off the mental property), but it does seem that, as a matter of fact in a case like Sosa's, the neural property does screen off the mental property. The worry then is that, if (II) is kept as a condition on causal irrelevance$_2$, then the causal irrelevance$_2$ of the mental will follow from AM after all.[18]

In response to this, notice first that (II)'s rendering the mental causally irrelevant$_2$ is independent of AM, at least to the extent that the problem-creating counterfactual, $(Nc \ \& \ -Mc) > Be$, holds whether or not there is a strict law linking N with B. So anyone who adopts (II) as a condition on causal irrelevance$_2$ will be committed to the causal irrelevance$_2$ of the mental in this case. But it seems to us that (II) is not a correct condition on irrelevance$_2$. It renders even properties connected by strict law causally irrelevant$_2$. To see this, consider the neural event c and the behavioral event e in Sosa's example. c possesses basic physical property P and mental property M (being a desire to quench his thirst), and e possesses the property B (being a certain movement of the hand). Assuming a strict law between P and B, it follows that:

(S) $(-Mc \ \& \ Pc) > Be$.

So, P screens off M from B. Now consider the counterfactual:

(T) $(-Pc \ \& \ Mc) > Be$.

It can be shown that (T) is compatible with AM *and* (S). Furthermore, it is plausible that (T) is in fact true. If c had been a desire to quench thirst but had not been P, it would have had some other property $P*$. Furthermore, c still would have resulted in an e that has the property B. That is, in the closest possible world in which Sosa desires to quench his thirst but this desire is not a P, it still causes him to extend his hand. Supporting this claim there may be a law, though not strict, to the effect that, when someone experiences a sudden

[18] Jerry Fodor has argued that a taxonomy of propositional attitude states in terms of their truth conditions is not a taxonomy in terms of causal powers. See his *Psychosemantics* (Cambridge, Mass.: MIT Press, 1987), ch. 2. Condition (III) may be involved in the view of some philosophers that scientific psychology requires a notion of narrow content. Thus, Fodor seems to hold that Oscar's belief that water quenches thirst is not causally relevant$_2$ to Oscar's behavior, since, if Oscar were in the same neural state as he is in but had not believed that water quenches his thirst, he would have behaved identically. The antecedent of this counterfactual is thought to be metaphysically possible for Putnamian reasons: if Oscar has lived in an environment containing XYZ and not H_2O his neural state would have been a belief that twin-water quenches thirst. One might conclude that, if we want a notion of content such that propositional attitudes are causally relevant$_2$ in virtue of their contents, then we need a notion of content which makes propositional attitudes supervene on neural states.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

desire to quench his thirst and believes there is a glass of water in front of him which he can reach by extending his hand, then, *ceteris paribus,* his hand will extend. When we consider the possibility that c is M but not P, this law "takes over" so that c still causes an event that is B. Here is a nonpsychological example which will, perhaps, help elucidate our claim.

Consider the event of hurricane Donald striking the coast causing the streets to be flooded. That event is identical to the event of certain air and water molecules moving in various complex ways. Call the property of consisting of molecules moving in such ways P. It is perfectly possible for the following counterfactual to be true: if hurricane Donald had not had property P (that is, if a hurricane as much like Donald as possible, though without P, had occurred), then it still would have caused the streets to be flooded. Indeed, it would have had some property $P*$ sufficiently similar to P, and $P*$ events (under the relevant conditions) cause floodings. The result is that Donald's being a hurricane would be said to be causally irrelevant to its flooding the streets. We think that examples such as this one show that (III) is too strong a requirement on causal relevance$_2$.[19]

A fully adequate account of causal relevance$_2$ should show how mentalistic counterfactuals are grounded. What is it about Sosa, his situation, etc., that makes it true that, if he had not experienced a sudden desire to quench his thirst, he would not have extended his hand? We do not have such an account, but we do want to suggest an approach that fits within the framework of AM. As we have observed already, the existence of nonstrict psychophysical and psychological laws is compatible with AM. A nonstrict law is one which has a *ceteris paribus* qualifier. The interesting thing about such laws is the ways in which they can support counterfactuals. We will illustrate this by building upon a suggestion by Lewis.[20] Let R, W, and B be the statements that a red block, a white block, and a blue block is placed in front of Donald and S_r, S_w, and S_b be the statements that Donald sees a red block, a white block, and a blue block. We will suppose, as is plausible, that there are nonstrict laws of the form:

(L) If X and C, then S_x,

[19] It may be that there is some account of causal relevance$_2$ midway in strength between (I) and (III) which captures what some of Davidson's critics have in mind. We leave to them the task of formulating it and attempting to demonstrate that AM entails the irrelevance$_2$ of the mental so characterized.

[20] "Causation," *Causation and Conditionals,* E. Sosa, ed. (Oxford, New York: 1975), pp. 180–191.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

where C are conditions like lighting is good, Donald is awake and paying attention, and so on. Even with such conditions added, the law is a *ceteris paribus* one and, if AM is correct, it will be impossible to add explicit conditions that turn it into a strict law. When the laws (L) hold, we will say that the statements describing what Donald sees *depend nomically* on the statements describing the blocks in front of him. Call conditions C *counterfactually independent* of the family of statements $\{R, W, B\}$, if C would continue to hold no matter which member of $\{R, W, B\}$ is true. Lewis shows that, if C and the *ceteris paribus* conditions associated with (L) are counterfactually independent of $\{R, W, B\}$, then S_x will depend counterfactually on X. That is, each of the counterfactuals, $R > S_r$, $W > S_w$, $B > S_b$, will be true. If we further assume that a block which has one of three colors will be placed in front of Donald [and that this statement is also counterfactually independent of (R, W, B)], then the statement $-X > -S_x$ will also be true. Suppose a red block is placed in front of Donald, and this event causes the event of his seeing a red block. It will follow that, if the first event had not been a placing of a red block, then the second event would not have been Donald's seeing a red block. As Lewis points out, this "grounding" of counterfactuals in laws fails to *reduce* counterfactuals to laws, since the assumption of counterfactual independence is essential. It does show, however, how laws, including *ceteris paribus* laws, can support counterfactuals. The program for a psychology compatible with AM is the discovery and the systematization of such nonstrict laws (at various levels) connecting psychological and/or behavioral properties.

We have seen that AM attempts to resolve the mind–body problem by endorsing (2), (3), and (3′), denying (1) with respect to events, and affirming (1) with respect to properties. Davidson is silent on (2) and (3) with respect to properties, leading to the accusation that AM is committed to epiphenomenalism. We rebutted this charge by showing that AM is compatible with there being counterfactual dependencies between events in virtue of their mental properties. To do this is to affirm (2) with respect to properties but, of course, to deny (3) with respect to properties. An event's physical features may counterfactually depend on another event's mental features. But, interestingly, we need not deny (3′) for our account of causal relevance$_2$. It may be that all counterfactuals *supervene* on basic physical truths and strict laws. That is, if two possible worlds are exactly alike with respect to basic physical facts and strict laws, they are exactly alike with respect to counterfactuals. This fairly strong physicalism still allows sufficient autonomy of the mental so that it is not reduci-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ble to the physical and it has a genuine explanatory and causal role to play.

ERNEST LE PORE

Rutgers University

BARRY LOEWER

University of South Carolina

MAKING MIND MATTER MORE*

Ernest LePore and Barry Loewer's paper argues that many of our intuitions about what is required for an intensional property of a mental event to be causally responsible for its behavioral effects can be squared with the metaphysics of "anomalous monism"—specifically, with the idea that singulary causal statements must be backed by strict laws (of physics). I argue, on the contrary, that anomalous monism really is incompatible with a robust construal of the causal responsibility of the mental and that some, at least, of its tenets will therefore have to be revised. In particular, I suggest, the requirement that singulary causal statements be backed by strict laws is unmotivated; all they require is backing by "hedged" laws whose *ceteris paribus* conditions are satisfied. This view not only legitimizes our intuitions about causal responsibility, but also provides a reasonable account of how the laws of the special sciences operate; or so I claim.

A consequence of this approach is that anomalous monism can no longer be invoked to underwrite the famous Davidsonian argument from mental causation to physicalism. An alternative argument is proposed which, though similar in spirit to Donald Davidson's, does not require the assumption that mental events are subsumed by physical laws.

JERRY FODOR

Graduate Center
City University of New York

* Abstract of a paper to be presented in an APA Symposium on Mind Matters, December 30, 1987, commenting on a paper by Ernest Le Pore and Barry Loewer, this JOURNAL, this issue, 630–642.

0022-362X/87/8411/0642$00.50 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

RELEVANT ALTERNATIVES AND DEMON SKEPTICISM*

THE degree of rigor we expect of ourselves and others in making claims to knowledge varies with context, and perhaps also with the speaker's intentions. This has seemed to many to provide, at long last, the basis for a definitive response to the skeptic. Roughly, the idea is that the skeptic attempts to impose extremely high standards of rigor, and that this yields two possibilities: either those standards are contextually inappropriate, in which case we may simply reject his claims that we do not know; or they are appropriate, in which case that very fact shows the context to be extremely unusual, even bizarre, and therefore of little concern—most of our claims are not made in such contexts, and are therefore not threatened. This perspective on central epistemological questions has become known as the *relevant alternatives* view. According to this view, again roughly, to know the truth of some proposition is to be justified in rejecting those alternatives which are relevant by contextually appropriate standards. I shall argue that the relevant alternativists are on to something important, but that, paradoxically, they themselves have failed to see what it is. My hypothesis is that the bogey of skepticism has pushed these philosophers into a serious distortion of their own central insight, but, whether or not this hypothesis is correct, there can be no doubt that something extraordinary has happened, since, given that insight, it is finally possible to see with real clarity that the skeptic is right! Or so I shall argue.

A closely and importantly related approach to the problem is associated primarily with Robert Nozick,[1] though it has precedents in the work of Fred Dretske[2] and others. The focus here is on showing that knowledge is not closed under known logical implication. That is, it is not true *in general* that, if S knows both that p, and that p implies q, then S knows that q. (In Nozick's view, all skeptical arguments rely on this principle, so that its refutation would suffice to refute all those arguments.)

But, although these positions are intimately related, their relations to skepticism and its possible refutation are quite distinct. This is a

* To be presented in an APA symposium of the same title, December 29, 1987. Jonathan Adler will comment; see this JOURNAL, this issue, 653/4.

[1] *Philosophical Explanations* (Cambridge, Mass.: Harvard, 1983).

[2] See "Epistemic Operators," this JOURNAL, LXVII, 24 (December 24, 1970): 1007–1023; "Contrastive Statements," *Philosophical Review*, LXXXI, 4 (October 1972): 411–430; "Conclusive Reasons," *Australasian Journal of Philosophy*, XLIX, 1 (May 1971): 1–22; and *Seeing and Knowing* (Chicago: University Press, 1969).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

result of their incorporating radically different conceptions of knowledge, which in turn are the objects of quite distinct skeptical challenges. I shall consider Nozick's view first, and then devote the bulk of the paper to the relevant alternativist position.

Nozick holds the following:

$$\begin{aligned} S \text{ knows } p \text{ iff (i) } & p, \\ \text{(ii) } & S \text{ believes } p, \\ \text{(iii) } & p \mathrel{\Box\!\!\rightarrow} S \text{ believes } p, \text{ and} \\ \text{(iv) } & \sim p \mathrel{\Box\!\!\rightarrow} \sim(S \text{ believes } p). \end{aligned}$$

(iii) and (iv) are subjunctive conditionals, which may be glossed as follows: (iii) is true just in case S believes p in all those possible worlds in which p holds true which are closest to the actual world, and similarly, *mutatis mutandis,* for (iv). Nozick argues that, given this conception of knowledge, one can show that the principle of closure is false. I know that I am now typing because: I am, I believe that I am, in the nearest possible worlds in which I am, I believe that I am, and, in the nearest possible worlds in which I am not, I do not believe that I am. I do not know, however, what I know to be implied by my typing, that *I am not being deceived by an evil demon into falsely believing that I am typing,* because even though that proposition is true, and I believe it, in the nearest possible worlds in which it is false, I still believe it, which violates condition (iv).

This argument simply begs the question against the skeptic.[3] Nozick, though he takes himself to be giving serious consideration to the skeptic's views by conceding that the evil demon hypothesis is possible, in fact *assumes* that it is possible *but not actual!* To say, in the relevant sense, that that hypothesis is possible, is to say not only that it is logically consistent, but also that *it may in fact be true.* But, if one concedes that, then one is in no position to determine what one's beliefs are like in those possible worlds closest to the actual one. Since the actual world may be evil-demon-governed (and may be any one of an infinite number of other possibilities), one simply has no idea of the truth or falsity of the subjunctive conditionals (iii) and (iv), whose truth is required for knowledge.

Relative to a Nozickean account of knowledge, the skeptic's position is that *we have no idea* whether we know our typical contingent beliefs to be true or not; that is, we have *no reason whatever* for

[3] I have argued this point somewhat more fully in "Nozick on Scepticism," *Philosophia,* XVI, 1 (April 1986): 65–69. In "Demon Scepticism," *American Philosophical Quarterly,* XIII, 3 (July 1986): 209–216, Doug Odegard repeats Nozick's mistake by taking it that the evil demon hypothesis is possible, *but false,* while purporting to provide an independent argument against it.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

thinking we have any such knowledge. Unless confused, he does not claim that we do not have such knowledge. How could he possibly be in a position to determine that? He has no idea what this world and its nearby neighbors are like, hence no idea whether the relevant subjunctives are true, hence no idea whether any of our beliefs constitute knowledge. This is, or should be, characteristic of the skeptic's position with respect to "externalist" views of knowledge, but not internalist ones. In spite of admirable sensitivity on the part of many philosophers to the distinction between skepticism with regard to the existence of (contingent) knowledge and skepticism with regard to our knowledge that we have such knowledge, it is an insufficiently noted consequence of the skeptic's skepticism that he cannot consistently take himself to know whether "external conditions" of knowledge obtain. Given an externalist conception of knowledge, he has no business denying that we *may* know a great deal; he can only point out, or argue, that we have no reason whatever to think we do.

The relevant alternativist's conception of knowledge differs sharply from Nozick's on this score. Relative to the former, knowledge of *p* requires that *I be able to justify* the rejection of all relevant alternatives to *p*. And here the skeptic's position will be that *I can be seen* to be in no position to do so; hence, *I can be seen* to lack knowledge. It is time now to turn our attention to this view.

I shall begin by considering an essay of Gail Stine's[4] which raises all of the right questions, while at the same time revealing crucial difficulties with the view. (The difficulties emerge more clearly in her work than elsewhere, principally because she worked out the position more completely than others have; these problems are inherent in the approach, as it has been understood thus far.) Stine develops her position in partial contrast to that of her fellow relevant alternativist, Dretske. Dretske, like Nozick after him, had argued against the closure of knowledge under known logical implication, but on classic relevant alternativist lines: we do not have any evidence on the basis of which to reject any number of hypotheses of varying degrees of wildness, but that is of no concern, since those hypotheses are not relevant alternatives to our actual beliefs. Thus, I know that the animals I see in the cage at the zoo are zebras, even though I have no

[4] "Scepticism, Relevant Alternatives, and Deductive Closure," *Philosophical Studies*, XXIX, 4 (April 1976): 249–261. Parenthetical page references are to this paper. Limitations of space prevent me from taking up here the earlier, closely related views of Ludwig Wittgenstein, especially in *On Certainty* (New York: Basil Blackwell, 1969); and of J. L. Austin, especially in "Other Minds," in his *Philosophical Papers* (New York: Oxford, 1961).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

grounds for rejecting the alternative hypothesis that they are mules painted to look like zebras; and I know that I am typing, even though I have no grounds for rejecting the alternative hypothesis that I am being deceived by an evil demon into falsely believing that I am.

Stine thinks that skeptical possibilities cannot be so flatly and definitively dismissed. According to her, the evil demon hypothesis *is* sometimes a relevant alternative in the context of a knowledge claim. The skeptic, as she puts it, has "an opening wedge"—"tighter criteria are appropriate in different contexts. It is one thing in a street encounter, another in a classroom, another in a law court—and who is to say it cannot be another in a philosophical discussion? . . . by some extreme standard, there is some reason to think there is an evil genius, after all" (254). The reason the skeptic fails to carry the day is that most knowledge claims are not made in philosophical discussions, hence the evil demon hypothesis is not *usually* a relevant alternative, and need not be ruled out: "a sceptic philosopher is wrong if he holds that *others* are wrong in any way . . . when they say they know a great deal"(*ibid.*). Those others are right, provided only that they have adequately ruled out the alternatives that are relevant in the contexts of their claims.

We must consider briefly the question of how the relevance of an alternative is determined, which is recognized on all hands to be a difficult problem. After faulting Dretske's proposal on the subject, Stine settles on the idea that an alternative is relevant if and only if there is some reason to think it is true. Thus, in discussing the question whether Henry knows that what he sees is a barn (in an "ordinary" situation she has described for illustrative purposes), Stine says that the possibility that it is a papier-mache fake is clearly relevant if there are such fakes in the immediate vicinity. It is less clearly relevant if there used to be such, but no longer are; or if there are none here, though there are some elsewhere; or if there are none now, though there used to be (here, or elsewhere). Stine is well aware that the matter of when there is some reason to think an alternative true is quite vague, but she thinks that for her purposes this is not problematic. The crucial point, as she sees it, is that there is *no reason whatever* to suppose that there is an evil demon; or so, at least, she holds when expounding this matter: "it seems safe to say [Descartes] was wrong if he thought there was some reason to think there *was* an evil genius. . . . the whole thrust of the relevant alternatives position as I conceive it, is that such an hypothesis is not relevant" (253).

But we have already seen Stine say that, "by some extreme standard, there is some reason to think there is an evil genius. . . ." What are we to make of this conflict? There is an air of self-referen-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tial paradox about any attempt to sort these matters out. A philosophical essay on skepticism is presumably a paradigm case of a philosophical context, and Stine herself has rhetorically asked the question, "Who shall say" what the appropriate standard is for such a context? Indeed, she has conceded that, by some (extreme) standard, the evil demon hypothesis is a relevant alternative, and it is clear that, on the skeptic's view, at least, it is precisely such a standard that is appropriate for such a discussion. Hence, in discussing with the skeptic whether Henry knows the object to be a barn, we must apparently take the evil demon hypothesis into account as a relevant alternative; we must, at any rate, if we are to be justified in claiming to take the skeptic seriously (as Stine does). Further, we must *conclude* that Henry does *not* know it is a barn, unless we can show that this alternative can be justifiably eliminated. It is presumably this situation that Stine has in mind when she says, "In truth, *in some sense,* scepticism is unanswerable" (254). The harsh truth seems to be that, if Stine is right in conceding the skeptic "an opening wedge," then we must ultimately conclude that the skeptic cannot be refuted *at all,* and not only "in some sense."

But, although she sometimes makes this apparently fatal concession, it appears to be her more settled view that *nothing at all* need be conceded, and here we find a second significant disagreement with Dretske. Although she goes to great lengths to defend the relevant alternatives position, with its aim of somehow isolating the skeptic, she *also* holds that the skeptic is *wrong* in claiming that his extreme standards cannot be met: she argues that *we know his radical alternatives to be false.* She takes Dretske to be right in thinking that we have no *evidence* that they are so, but wrong to abandon deductive closure on those grounds: we know they are false, because "one does know what one takes for granted in normal circumstances. . . . The evidence picture of knowledge has been carried too far" (258). If this last point is right, then we simply have no reason not to adopt the skeptic's standards: we can cheerfully admit his challenges as legitimate, since they are adequately met by the fact that we take the falsity of his outrageous alternatives for granted, and hence know them to be false.

To my knowledge, others have not openly followed Stine in this last, somewhat desperate proposal. One reason is perhaps obvious enough: in the history of the race, any number of false views have been taken for granted by any number of people. But, if we do not follow her here, we are left with the tensions that, I surmise, originally motivated the proposal, to wit, the inconsistency of her other views: (i) there is no reason whatever to suppose that there is an evil

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

demon; (ii) by some standard (or in some contexts) the existence of a demon is a relevant alternative; and (iii) an alternative is relevant if and only if there is some reason to think it true.

Before turning to work these matters out, I want briefly to consider a recent effort to provide some independent support for relevant alternativism. Stewart Cohen[5] has argued that whether an individual possesses knowledge is a function of the intellectual capacities of a relevant epistemic community: *S* knows *p* provided *S* truly believes *p* and *S* possesses no defeater of his justification for *p*, whose status as a defeater is intersubjectively obvious within that community. In his view, it is possible for a speaker to determine deliberately which community, with its accompanying standards of obviousness, is the relevant one, simply by intending to use its standards, though "routinely our own personal/social perspective determines which standards we intend" (580). This has the consequence that I can correctly claim to know *p*, while insisting that a member of a genius society possessed of exactly the evidence available to me may not know *p*, since, by the standards of his society, that evidence may include a defeater whose status as a defeater is intersubjectively obvious in the genius community. In the other direction, I could attribute knowledge to a member of a moron society by intending my claim to be evaluated by the standards appropriate to that society, while simultaneously denying that I, who possess exactly the same evidence, have knowledge, since by my standards there is a defeater whose status as a defeater is intersubjectively obvious to the members of my society. This conclusion is reached independently of any considerations of skepticism, and Cohen argues that this provides support for the relevant alternativists' parallel response to skepticism, to wit, the claim that *S* knows *p* if and only if *S* truly believes *p* and is in a position to rule out all the relevant alternatives to *p*, where the relevance of an alternative is determined by its objective probability.

My argument against all of this will consist primarily of my positive account, but several things should be noted in passing. First, Cohen's theory of knowledge, developed independently of skeptical considerations, carries a very high price—it makes knowledge a negative function of intellectual capacity—and this makes it quite unpromising as the basis for an independent defense of the relevant alternatives position. Second, Cohen shares in the widespread confusion concerning how an alternative's relevance is determined; he indicates, first, that the mere presence of papier-mache barn replicas in

[5] "Knowledge and Context," this JOURNAL, LXXXIII, 10 (October 1986): 574–583.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the vicinity makes relevant the alternative that what *S* sees is a barn replica, but he subsequently suggests that more is needed, e.g., knowledge that a reliable source reports the existence of the replicas. Finally, Cohen clearly implies that there are contexts in which the evil demon hypothesis is a relevant alternative, or in which we are free to regard it as such, while maintaining also that the relevance of an alternative is a matter of the probabilistic standards in effect. But, I shall argue, Stine is correct in thinking that there is absolutely no reason to believe that there is an evil demon, and by that I mean that there is no available evidence, there are no available data, relative to which that hypothesis has any probability at all. The relevance of that alternative has nothing to do with its having any finite probability relative to anything.

It is time to try to separate the relevant alternativists' insights from their confusions, and to see to what extent the insights can be preserved while drawing out the skeptical consequences I take to be implicit in their view. In "Scorekeeping in a Language Game,"[6] David Lewis has provided an elegant account of how the presuppositions operative in a conversation can shift with the course of that conversation: "If at time *t* something is said that requires presupposition *P* to be acceptable, and if *P* is not presupposed just before *t*, then—*ceteris paribus* and within certain limits—presupposition *P* comes into existence at *t*" (234). One of his illustrations of this phenomenon is an imagined dispute with a skeptic over the appropriate standards for the use of such vague terms as 'flat' and 'certain'. Here the idea is that the presupposed standard of precision for use of these terms varies with the context, and that, in the course of a conversation with the skeptic, the standard gets raised extremely high—perhaps to the point that the term has no application at all. Lewis's point is that that turn of events has no relevance to the truth of claims made earlier in the conversation, or in other conversations, when the standard was lower.

Imagine now a discussion following a presentation of Lewis's paper, in the course of which an innocent who happens to be present asks Lewis: "Well, are you saying that billiard tables (in good condition) are flat, or not?" or simply, "Are billiard tables (in good condition) flat?" This brings out vividly a central Lewisian point—the answer to such a question is determinate only in a conversation in which some relevant presupposition or other exists. The innocent's question *has no answer*, precisely because no such presuppositions are in place. But the flatness of billiard tables is not of much interest;

[6] In *Philosophical Papers*, vol. 1 (New York: Oxford, 1983).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

we must now try to understand the implications of this view for questions concerning human knowledge.

I am much taken with Thomas Nagel's[7] idea that often, in grappling with philosophical questions, we are trying to reconcile two views of the world, one subjective and one objective. It seems to me that the relevant alternativists are doing just that. They see that, in any concrete conversation, we are guided in what we say by certain relatively clear presuppositions, which may or may not govern other conversations, and, seeing this, they hope to portray the skeptic as either (i) failing to understand this point, and trying to impose a particularly rigid set of presuppositions on every conversation in which anyone's knowledge of anything comes in for mention, or (ii) making claims which are true (since his presence and activity determine the relevant presuppositions), but which have no bearing on ordinary claims to knowledge which appear superficially to be in conflict with them.

But all of this is more or less on the subjective side of the ledger. It emphasizes the legitimacy of our attending to the practical aspects of our commerce with the world, and that is all to the good, perhaps. Its failure lies in giving inadequate recognition to our pull toward objectivity, as reflected in another question we might have put in the mouth of our innocent: "Do we know there's a physical world we inhabit, or not?" It is just this tension, I suggest, that leads Stine to say both that there is no reason to think there is an evil demon, and that by some extreme standard there is after all some reason to think there is an evil demon.

Let me say, then, how I think sense is to be made of all of this. So far as I know, no relevant alternativist has noticed that the evil demon hypothesis [that is, Descartes's classic one, according to which nothing exists or ever has existed but an evil demon and my disembodied mind with its myriad (subjective) states] cannot be a relevant alternative except in one very special context, or at any rate one special sort of context, namely, one in which *I presuppose nothing except the existence of my own mental states*. The reason for this is that the existence of anything other than those states is *logically incompatible* with the evil demon hypothesis! Consequently, in *any* context in which *any* other existential presuppositions are in place, the skeptic cannot make the demon hypothesis a relevant alternative by raising standards of *probability* so high that even that vastly "improbable" hypothesis cannot be ruled out. It is *completely* ruled out, on logical grounds alone, by the presence of such presuppositions.

[7] *The View from Nowhere* (New York: Oxford, 1986).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

But there are, I have just said, special contexts in which that hypothesis is a relevant alternative. What is its fate in such contexts? Even there, as Stine sometimes sees, *there is no reason whatever* to suppose that such a demon exists. But what has seldom if ever been noticed is that *neither is there any reason to think there is a physical universe* which we inhabit. Relative only to the existence of my subjective states, there is an infinity of hypotheses about what else reality may comprise, and each of them has the same standing: there is no reason whatever to accept any of them. The relevance of the demon hypothesis stems not from the application of some extreme standard of probability, but from its membership in an infinite set of hypotheses, all of which are equally and completely devoid of support.

The situation is quite analogous to the one in which Pascal found himself: taking for granted the existence of a physical universe, he recognized that he was "infinitely ignorant" with respect to the question of God's existence. (Notoriously, he failed to see the full force of his own argument, but that is another matter.)

Elsewhere[8] I have argued, somewhat unfashionably, that my (largely unspoken and unconscious) beliefs about how things seem to me to be and to have been hold a privileged place in my larger body of beliefs—they are, as I have put it, forced on me (in just the way that philosophers since the time of the Greek skeptics have tried to say they are). It is a coherent possibility, no matter how absurd, that I am not now sitting at a typewriter; but *for me* it is not a coherent possibility that it does not *seem* to me that I am. Thus, I suggest that the fact that it seems to me that things are a certain way, and not others, is an *unavoidable* presupposition in each of my conversations. (But notice that it is *my* presupposition, and not *ours*.) The skeptic, therefore, has a strong basis for claiming that the context which interests him is of special importance—it is the one in which I make no presuppositions except those which are utterly unavoidable. And, in that context, he claims, it is clear that *either* I have absolutely no knowledge of contingent matters of fact beyond those which I presuppose (in case the concept of knowledge is to be cashed out in terms of justification), *or* I have absolutely no reason to believe that I have any such knowledge (in case the concept of knowledge is to be cashed out in terms of reliable processes, subjunctive conditionals analyzable in terms of possible worlds, etc.).

But what, then, of other contexts? Is the skeptic prepared to agree

[8] "The Given," *Philosophy and Phenomenological Research*, XLVI, 4 (June 1986): 597–613.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

with the relevant alternativists that there need be nothing wrong with knowledge claims made there? Yes; but that is less of a concession than might be supposed. Stine was wrong to claim that we know whatever we take for granted, but there is an important truth in the neighborhood: we *take ourselves* to know what we take for granted, and against that background we take ourselves to be seeking further knowledge. And now we are in a position to make good sense of the relevant alternativists' principal insight. The crucial point is that I take for granted that you, as I, continuously take for granted a four-dimensional world which is our common home, and indeed that we both continuously take for granted a good deal more—that our individual and collective powers of observation and memory are more or less reliable, and so forth. Thus, I take for granted both that we mutually take for granted that we continuously share this vast range of things taken for granted, and that we have devised a way of speaking which is consonant with this situation: we attribute knowledge to each other and to ourselves provided that we can see that those alternatives to our beliefs which are relevant, given what we mutually take for granted, can be ruled out. (Incidentally, this perspective on knowledge helps make clear that the Gettier problem and the problem of skepticism about "the external world" are entirely independent of one another, so that recent attempts to solve both in one grand sweep were always doomed to failure. The Gettier problem is a baffling problem about how to assess knowledge claims in contexts in which much is taken for granted; a solution to that problem would have no bearing whatever on the skeptic's claim that, relative only to the unavoidable presupposition of our knowledge of our own conscious mental states, we know nothing at all about the existence of an external world.)

Philosophical reflection is a late arrival on the human scene—certainly arriving much later than the development of a more or less comprehensive language. The skeptic's contribution to our understanding of ourselves, when finally and properly understood, consists in pointing out that, at best, we have no reason to think we know the vast number of things we continuously take for granted. But once it is recognized on all hands that our knowledge claims are (almost) universally to be taken as relative to some body of presuppositions, and that we have no reason to think we know our deepest optional presuppositions to be true, since they are entirely unsupported by our unavoidable presuppositions, it would be both mistaken and churlish of the skeptic to object to our continuing to speak in the ways we always have. He should be satisfied to have us concede, in some appropriate context, that the evil demon hypothesis is no more

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

outrageous than our own favored hypothesis, and he should then be prepared to let us get on with our normal practice of taking for granted our familiar world.

BREDO C. JOHNSEN

University of Houston

RELEVANT ALTERNATIVES, PRESUPPOSITIONS, AND SKEPTICISM*

Instead of attempting to refute or show incoherent the skeptic's doubts, Bredo C. Johnsen adopts the strategy of "boxing-in." He attempts to show that the skeptic's hypothesis is viable only in very special contexts. Admitting it there does not threaten to spread, so as to infect our everyday empirical claims in their normal contexts. The result of this strategy is a reversal of the burden of proof. Similarly, the relevant alternatives' approach attempts to offer a neutral account of how we normally delineate the range of counterpossibilities which the putative knower must address. The skeptic's hypothesis is outside this range, so the burden is on the skeptic to demonstrate its relevance.

Johnsen is on firmer ground than many relevant alternative theorists in recognizing the extreme sensitivity of the determination of relevant alternatives to contextual matters. Although this recognition diminishes the burden forced upon the skeptic, Johnsen makes the burden much greater in other ways. For he seems to conceive the evil demon hypothesis as primarily an ontological hypothesis—one claiming that nothing ever existed except the evil demon and my disembodied mind.

A different view of skepticism, and of the location of the burden of proof, emerges if we approach matters from a basic Cartesian perspective which asks us to determine the proper epistemological standards as if our interest were solely in attaining truth. This approach provides a rationale for radical skepticism which unites it with skeptical concerns generally, rather than being restricted to the singular

* Abstract of a paper to presented in an APA symposium on Relevant Alternatives and Demon Skepticism, December 29, 1987, commenting on a paper by Bredo C. Johnsen; see this JOURNAL, this issue, 643–653.

0022-362X/87/8411/0653$00.50 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

case of the evil demon. In the terms Johnsen adopts, this perspective is more objective.

Johnsen observes that the relevant alternativist over emphasizes the subjective side of knowledge at the expense of its objective side. I agree. In particular, I shall try to show that our basis for delineating the class of relevant alternatives is itself motivated by prudential considerations to avoid radical skeptical consequences and our (practically necessary) predilection for recalling and recording our empirical successes, rather than our failures.

I wonder, though, if Johnsen too does not share a bias toward the subjective side. There is much value in Johnsen's emphasis on presuppositions, what is taken for granted, and contextual standards. Among other insights, it helps explain, and often explain away, some of the initial resistance to, and then the appeal of, skeptical arguments. Context independent views or, at least, views that demand very high contextual standards for knowledge remain viable, however. Aside from the objective approach noted above, I will try to relate Johnsen's concerns to the views of Peter Unger in his recent *Philosophical Relativity*.[1] Unger, like Johnsen, is influenced by David Lewis's important work. But Unger tries to show us that we can maintain a completely noncontextualist approach—what he calls "invariantism"—that welcomes radical skeptical conclusions. Unger argues that we have no rational basis for deciding between invariantism and contextualism.

JONATHAN E. ADLER

Brooklyn College, City University of New York

[1] (Minneapolis, Minn.: Minnesota UP, 1984).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

PUBLIC POLICY AND THE ALLOCATION OF SCARCE MEDICAL RESOURCES*

WHEN available medical resources are insufficient, as they so often are, to meet the needs or desires of those who seek them, what principles or factors should determine their allocation? If a society responds to such an observed or perceived insufficiency by deciding to produce or acquire additional resources, or to re-allocate existing resources, sufficient to meet existing or predictable needs and claims in a given area, what medical or nonmedical factors or principles justify such a decision? If a society decides that its macroallocations of general economic resources to the providing or paying for the costs of medical interventions are already excessive, and so takes actions designed to diminish these, what medical, ethical, or economic principles might rightly be invoked to support or to oppose such actions? Do these questions change significantly with the degrees of insufficiency, ranging from "almost enough" to "critically short" to "rare" and (therefore probably) experimental? Because of the different needs for treatment, are they different for emergencies, acute and chronic illnesses, a single disease or dysfunction, the whole of health care, preventive measures, radical therapeutic interventions, or for life-saving, restorative, palliative, or elective procedures?

I.

Many answers to these questions have been proposed. Their range and variety are impressive. Not only does the pluralism of our culture issue in a great number of differing proposals and positions; on the world scale, even greater diversity emerges, and, within the major philosophical traditions, even some which seem fairly compact and homogeneous, we find significant variations.

There are those who hold every human being and life to be of absolute and infinite—and therefore equal—value; differential calculations and specific microallocational decisions would thus be ethically—and even logically—unacceptable. Priorities or choice could only be properly established among these infinite and equal values by external factors such as random selection or a first-come first-served policy. Anything less arbitrary would derogate the dignity not only of those denied treatment, but even of those favorably selected. According to an even more absolutist view, either all of those who need a given resource should receive it, or else none should.

* To be presented in an APA symposium of the same title, December 29, 1987. Norman Daniels will comment; see this JOURNAL, this issue, 664/5.

0022-362X/87/8411/0655$00.90 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

At a far remove from this view, which makes of medical intervention such a special matter because of the unique value of its beneficiaries, are those who hold that the benefits of medical care are in themselves not very different from all the other goods and services subject to the economics of scarcity which are sought, produced, and distributed by human beings for each other. They therefore propose that all such allocational determinations be made by and in the marketplace, where the autonomous wills of the providers and consumers, insofar as these are enabled and sustained by their competencies and purchasing powers, will drive and govern both macro- and microallocational processes. Minimalist states will provide only an orderly structure within which such transactions may take place.

Others hold the society, not the individual, to be of ultimate concern, even in the doctor–patient relationship; they therefore urge that the state has the right and duty, and should have and exercise the power, to make all such decisions, and to make them through its political processes in such a way as to benefit the society as a whole most greatly or to harm it the least, so far as these outcomes can be foreseen. Only such individualism would be tolerated as would itself contribute or conform to the greatest benefit/least harm to society principle. This stricture would also apply to other traditionally developed concepts in medical ethics, such as confidentiality, informed consent, and assessment of risk/benefit and cost/benefit ratios, and of social worth or desert.

Still others, while not maintaining the absolute or ultimate value of each life nor even the equality of all, do hold the one-to-one doctor–patient relationship to be unique or at least very special among human transactions, with a mixture of rights and duties derived autogenously from the contract, covenant, or commitments that have been made; once established, this relationship and its inherent rights and duties have such moral autonomy as to resist rightly any external pressures, at least for all microallocative decisions involving primarily the two principals. Some who hold this view would prefer that macroallocative decisions and policies be, so far as possible, the sums or aggregates of individual ones; others seem to allow them to be produced by the normal social and political processes embodying and exemplifying whatever may be that society's conception of distributive justice.

Although not holding a hard-line professionally deontological position, there are yet others who see the acts of medical intervention as special ones, with motivations, purposes, trusts, and end products to be significantly distinguished from those of all or most other human

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

transactions. Although such traditional concepts of ethics as autonomy, beneficience, and justice would apply to medical care, they would do so in quite special ways. The allocational policies and decisions implied by these special applications are not easy for me to characterize because of their diversity of detail, but they round out my case to this point, which is that the answers proposed to my original question are numerous and diverse, and increasingly so as the literature grows. Is there a reasonable way to deal with this diversity, other than by taking a partisan stance in favor of some chosen theory and against all those which appear to differ from it or at least to contravene it? I am led to hope that there is, and will now present a brief account of it.

II.

So much has been written, by so many and so well and persuasively, on the subject of medical resource allocation and its ethical aspects and problems. A fair question at the outset of this paper might be: Why write more? My initial and modest answer is that I will attempt to focus attention on a few critical concepts, used frequently but by no means univocally, and thereby sharpen or clarify some of the issues. I will also attempt, somewhat less modestly, to find a way of coming to terms with the considerable diversity of views encountered in these sources.

Let me first note, with gratitude, some of the thinkers who have written so helpfully and ably. I am very much in their debt for the wide range of speculations, insights, and contentions they offer. Three recent books have contributed much of my understanding and interest; each deals with allocational problems within the context of a broad and coherent philosophical tradition. Edmund D. Pellegrino and David C. Thomasma[1] address these issues within a realistic and traditional context, strong in its Aristotelian roots but eclectic and pluralistic in its analyses of contemporary problems. Robert M. Veatch[2] offers a systematic and original application of an evolving social contract theory to a wide range of issues and cases. He deals more broadly with the central concern of this essay though more by implication than by specific and explicit treatment, but I think these implications are reasonable and clear. At the conclusion of the book he offers "A Draft Medical Ethical Covenant" which strongly endorses equality of access to health care insofar as this is necessary to

[1] *A Philosophical Basis of Medical Practice: Toward a Philosophy and Ethic of the Healing Professions* (New York: Oxford, 1982).
[2] *A Theory of Medical Ethics* (New York: Basic Books, 1981).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

provide opportunity to achieve equal levels of health among recipients, so far as this may be possible. H. Tristram Engelhardt, Jr.[3] essays a comprehensive theory based upon "rational authority" in a world characterized by the continuing existence of a large and often unyielding number of moral viewpoints and theoretical systems, including many strongly deontological and religious ones. His work has been seen by some as a defense or attempted justification of secular pluralism in morality, but he foresaw and specifically rejected this assessment.

A fourth book, by Albert R. Jonsen, Mark Siegler, and William J. Winslade,[4] has also been very useful to me because of its sharply focused emphases on those microallocational decisions which must be made at or very close to the bedside; for it is here that I wish to concentrate some of my attention. Also, two modern classics in moral philosophy must be noted because of the frequency of their citations throughout much of the literature on this subject, and the influence their conceptions of justice have had on so many writers. These are, of course, *A Theory of Justice* by John Rawls,[5] and *Anarchy, State, and Utopia* by Robert Nozick.[6] I will not attempt to give similar citations to all of the articles and monographs which have made some contributions to my reflections on this matter; not all of them focus directly on this problem, but I do feel that a reference to some authors[7] may be helpful to the reader who wishes to pursue the question, and may also clarify the intent of the position I reach on the treatment of diversity and comprehensiveness.

To all these and to many others I owe much gratitude for their individual explications and for their collective range and diversity. A search for consensus internally or by abstraction is clearly doomed to failure, even more so as the list of sources grows longer. While often citing and praising one another's work, they show very little movement toward any common core of agreement. This worries and intrigues me, since I respect the work of each. Thus, I am led to ask

[3] *The Foundations of Bioethics* (New York: Oxford, 1986). Cf. two perceptive, but differing critical views of Engelhardt's work by David H. Smith and Gregory W. Trianowsky in *Medical Humanities Review*, I, 1 (January 1987): 11–14 and 15–20, respectively.

[4] *Clinical Ethics*, 2nd ed. (New York: Macmillan, 1986).

[5] (Cambridge, Mass.: Harvard, 1971).

[6] (New York: Basic Books, 1974).

[7] Dan Beauchamp, Tom Beauchamp, Edmond Cahn, Daniel Callahan, Arthur Caplan, Alexander Capron, James Childress, Norman Daniels, Charles Fried, Ruth Faden, Stuart Hampshire, Thomas King, Lawrence McCullogh, Paul Ramsey, Nicholas Rescher, Robert Sade, and LeRoy Walters.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

whether it may be possible to address this great variety of serious thinkers and their views by endorsing their affirmations and acknowledging the reality of the values they promote and the justifications they present, while not being bound by their denials or exclusions, nor overly dismayed by what clearly appear to be contradictions or incompatibilities among their very affirmations. I am led to this approach by considerations at two levels: medical practice and the history of moral philosophy.

Microallocation decisions, at both of Engelhardt's levels, are being made every day by physicians and those who work very closely with them; for the great majority of these, and quite possibly for all of them, unless they are flawed by some glaring error of knowledge or intent, a justification could be found among the cases, exemplifications, or principles in the literature. From the other direction, for every value or principle to be found in the literature of medical ethics on the subject of the microallocation of limited resources, I believe cases or instances can be found in the practice and literature of medicine. This is not to claim that each decision was absolutely or inherently right and good, nor even that it was the best among the alternatives, but rather that a reasonable justification could be found for each decision. I do, of course, mean a moral justification, which requires that some reasonable persons could be persuaded that significant human values were part of the intended outcome or result of the decision, and were part of the reason for making it.

At the macroallocational levels, reasonable and conscientious decisions are also being made by a wider and more diverse group of persons. Again, I would propose that, for each decision, a reasonable justification can be found in print or public policy, or soon to be there. Again, no claim of absolute or ultimate rightness and goodness is being made, only one of reasonable justifiability. If these contentions are correct, is it not then reasonable to say that, in allocational decisions at all such levels and for all degrees of scarcity of resources or overages of established needs, every value that is at issue or under consideration is worthy of such attention, and that the most desirable end result of such considerations is the decision that will maximize the sum or synthesis of these values in a real or pragmatic composite. Is it not also reasonable to say that, as each of the theoretical positions set forth has its own rationale and its probable exemplifications in justifiable decisions, so each is tenable, subject to the limitations imposed by the understanding by each that each of the others is also tenable; the ultimate theory can then be defined, though not yet formulated, as the synthesizing and maximizing of all

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the rightness claims and goodness claims of all the extant theories, but with an open-ended readiness to encounter and consider all serious theoretical claims yet to be made. Some may see this as a failure to decide among competing values and competing theories. I propose instead that it is the most difficult and ambitious of decisions: the decision to include and consider respectfully every value claim and every theory claim, and then to try to make the most comprehensive and maximizing synthesis possible.

A good example of this kind of multivalued maximizing decision making, literally or metaphorically at the bedside, is to be found in a compact article by Dr. Thomas C. King.[8] Addressing the ethical problems surrounding or permeating patient care in an intensive care unit, he considers the broad categories of admitting decisions and policies, patient management, ethical conflicts within the team, financial consequences, and risk/benefit analyses; several of these have a number of subcategories. Noting that, "The ICU is a microcosm of dilemmas facing the entire medical care system" (174), King states that the clinician must weigh and balance as many of the values and possibilities of outcome as can be discerned, make a pragmatic decision on the basis of limited knowledge of the probabilities and some clinical guesses, and hope for the best outcome, medically and ethically. I defer to, and concur in the results of, his experience and judgment. Good, or at least justifiable, decisions are being made in just this way in ICU's all over the country and in their counterparts throughout much of the world. Public policy decisions guiding macroallocation are subject to the same logic; the persons and institutions generating them are more complex and diverse; the values to be considered include all relevant individual and social results or outcomes of the medical interventions, as well as those inherent in the clinical processes themselves, and this on the large scale; but the obligations of the decision makers are still to maximize the values of all involved, so far as these can be discerned and effected. Although the practical demands of this approach are heavy and the difficulties of selecting the best course of action, or even knowing retrospectively that you have done so, are clear and admitted, the theory is simplicity itself.

III.

One's position on a specific topic or issue in medical ethics should be coherent with one's overall views of that discipline as a whole, of

[8] "Ethical Dilemmas of Restricted Resources," in *Human and Ethical Issues in the Surgical Care of Patients with Life-Threatening Disease*, F. P. Herter et alia, eds. (Springfield, Ill.: Thomas, 1986), pp. 169–175.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ethical theory, and of philosophy itself.[9] It will try to show this coherence.

Medical interventions aim primarily at health, physical and mental well-being, and maintenance of life. Such values as confidentiality, trust, respect, compassion, truthfulness, beneficence, freedom from pain, and many others familiar to us are frequently in harmony with those primary values and their means, and are often instrumental to those primary values or corollary effects of their achievement. But conflicts among the corollary values, and between some of them and the primary values, are often practical realities with which clinicians and policy makers must deal. A major purpose of the discipline of medical ethics is to discover ways of maximizing the sum or synthesis of all of these values, which may be mutually enriching, competing, or contradictory in some cases. The divergence or actually contradictory status of some values in real situations calling for decision should not deter the ethicist from the search for the maximizing solution, nor lead anyone to claim that the original values were not what they were originally seen and felt to be.

I look with great respect upon the history of moral philosophy with all its diversity and multiplicity of ethical theories. Each attempt to discover, understand, and express an account of human values and rules of conduct, our duties, rights, and appropriate benefits, and the nature of moral rightness and goodness seems clearly to be worthy of my consideration. Those which have survived well the effects of time and critical analyses seem especially deserving. At the base of all these theoretical speculations and contentions lies what may be called descriptive axiology, which I also view with respect. Here are the historical, psychological, sociological, and cross-cultural accounts of all the values held, pursued, admired, or enjoyed by human beings. This constitutes an important part of the data of ethics, which merits the initial consideration of the moral philosopher; although none of us can ever be aware of the full range of this data, we can in principle be open to it and aware of its open-ended nature. All such value claims or value reports deserve the consideration of the ethical theorist on two levels. First, on the descriptive and explanatory level, and, second, on the normative or critical. Each ethical theory should be able to demonstrate its awareness of this domain of values; it should then be able to include any given value

[9] Cf. C. Danner Clouser, "What Medical Ethics Is," in *Annals of Internal Medicine,* LXXX, 5 (1974): 657–660. "Medical ethics specializes in applying the old familiar rules to one particular domain of activity, namely, the medical."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

claim rationally within its theoretical structure, or to exclude or transvalue it in such a way that reasonable persons could understand the exclusion or modification. The exclusion will not be a denial of the original value claim, but an ordering in which a lesser or lower value must give way to a higher or greater one, or be left out of a maximizing synthesis as a result of conflict resolution.

If each of the enduring moral philosophies can claim some degree of success in these matters, and I think each can, then we are faced with a sizeable number of tenable theories, each earnestly offering to tell us how life should be lived, how conduct should be governed, how decisions should be made. But values do conflict and, in the real world, choices must be made: we strive toward some values and forego others. The rational key to such choices is maximization, the putting aside of some values only because a greater value or sum of values is thereby attained. The best course of action is the one which attains the greatest sum of values. So too when we choose between ethical theories that differ and compete for our assent, the key will be that theoretical synthesis which best harmonizes and maximizes the aims and goals of the several theories considered, and which is most open to the consideration of theoretical claims yet to be made.

A similar approach to philosophy is indicated. The ability to comprehend and synthesize the even greater diversity of philosophical systems, components, and fragments of insight, to account for those, and to show how they are to be understood in the context of the synthesis will be the hallmark of this philosophy of philosophy.[10]

IV.

I conclude with a note on medical needs. In the work cited above, Pellegrino and Thomasma, characterizing medicine as a moral enterprise, state that its goal is "its central aim; a right and good healing decision taken in the interest of an individual patient or society in need" (224). I wish to assert that need is axiomatic in establishing eligibility for resource allocation. Norman Daniels[11] has written instructively about health care needs, and I urge attention to his treatment to establish clearly the differences between his more expansive meaning and my own limited one. In my view, a medical need exists when a person suffers some malady that significantly diminishes or threatens what that person regards as normal health or functional

[10] I have written briefly on this theme elsewhere. "Philosophic Disagreement and the Study of Philosophy," in *Tulane Studies in Philosophy*, VII (1958).

[11] "Health Care Needs and Distributive Justice," in *Philosophy & Public Affairs*, X, 2 (Spring 1981): 147–179; also see chs. 2–3 of his *Just Health Care* (New York: Cambridge, 1985).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

capacity, and (that person) believes that therapeutic interventions by skilled and caring medical professionals might cure or alleviate that malady (injury, illness, dysfunction, pain, incapacitation, etc.). These elements establish a subjective medical need. If a consensus can be reached regarding the existence of the malady or of some underlying condition of which the malady is an effect or symptom, and if the consensus includes at least one qualified and reasonable observer, preferably a medical professional, the need can be called quasi objective. This account must be supplemented by two limited cases: one in which the qualified professional engages in competent self-diagnosis; the other in which the person believed to have the malady lacks autonomy or self-awareness, being, for example, an infant, mentally ill or comatose, or otherwise unperceptive of the malady and unable to believe in the efficacy of therapeutic intervention. In all these cases, if the consensus includes several qualified observers, and all or most concur in the diagnosis and in the hope that intervention may be of benefit, I would term the focus of these perceptions a fully objective medical need. Persons having such needs are eligible for the microallocation of scarce medical resources. If priorities must still be established within this group of eligible persons, I suggest consideration of the following probabilities: (1) that intervention may presently make the difference between life and death; (2) that intervention may make a major difference in the health and vital capacities and therefore in the length and quality of life; and (3) that intervention may enhance the quality of life, but probably not its length. Since need has been presented as a function of both the presence of malady and the prospect of helpful intervention, such considerations, combining as they do some measure of the gravity of the former and the probabilities of the latter, seem appropriate ways of quantifying or rank ordering the needs differentially and therefore the allocational priorities.

RICHARD L. BARBER

University of Louisville

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

JUSTICE AND THE DISSEMINATION OF "BIG-TICKET" TECHNOLOGIES*

In his "Public Policy and the Allocation of Scarce Medical Resources," Richard Barber claims that conflicting "values" underlie disputes about how scarce medical resources should be allocated, and then proposes his "solution"—that we "maximize values" in order to produce the "best outcome." He fails to distinguish macro and micro decision-making contexts or the different principles or procedures appropriate to each; he tells us nothing about what values are, about how we should quantify or "weigh and balance" them, or about how to "maximize" them when they conflict; he offers no arguments about substantive issues; and he does not apply his solution. In short, his discussion does not engage (and is not representative of) the rigorous philosophical literature on this topic.

Rather than tilt at Barber's windmill, this Don Quixote will poke at one of his own. I shall focus on the more specific question: Why are health-care institutions so unresponsive to certain arguments about the just distribution of some high-cost medical technologies? When we are concerned with technologies whose total costs may add up to several billion dollars a year, it is reasonable to ask about their opportunity costs: What else could be accomplished with the same investment? For example, organ transplantation shares an important property with a number of other high-cost medical technologies: they all deliver a significant benefit to a small number of people at a high cost. This fact makes it seem likely that the opportunity costs of these technologies will be problematic—more could be accomplished with alternative investments.

Arguments based on such claims about opportunity costs can be recast into arguments about just distribution. I consider how such arguments might look on both utilitarian and "fair equality of opportunity" accounts.[1] My main concern is to show that certain features of health-care institutions in the United States make it especially difficult for decision makers to respond to opportunity-cost arguments, even when they are based on claims of justice. One type of

* Abstract of a paper to be presented in an APA symposium on Public Policy and the Allocation of Scarce Medical Resources, December 29, 1987, commenting on a paper by Richard L. Barber; see this JOURNAL, this issue, 655–663.

[1] See my *Just Health Care* (New York: Cambridge, 1985), and *Am I My Parents' Keeper? An Essay on Justice Between the Young and the Old* (New York: Oxford, forthcoming).

0022-362X/87/8411/0664$00.50 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

problem is quite general and not specific to our health care system: arguments based on claims about *ideal* justice are problematic guides to decision making in partial compliance and noncompliance contexts. Another set of obstacles facing opportunity-cost arguments is traceable to specific features of our health-care system. Most of our health-care decision making is not regionally or nationally coordinated. Different providers often directly compete with each other, which gives them some different, merely agent-relative, rather than common goals. Even where there is no direct competition, there is still no coordination and cooperation. When providers *successfully* pursue their agent-relative goals, they may still produce outcomes that are collectively worse for them.

I shall describe how a type of many-person contributers' dilemma arises in the context of providers' deciding how to disseminate new medical technologies. The dilemma is similar to Derek Parfit's Parents' Dilemma,[2] and it shows that the difficulty we face in medical contexts is not just the result of narrowly self-interested motives biasing allocation decisions. Other-regarding "agency" motives, e.g., doing the best we can for our patients, can give rise to the same pattern of failing to respond to opportunity-cost arguments that appeal to justice, with the result that too much high-cost technology is disseminated. In this case, the solution is not the promotion of more altruistic behavior on the part of decision makers, but the imposition of regulations that reflect social goals and priorities involved in the just distribution of health-care resources.

NORMAN DANIELS

Tufts University

[2] *Reasons and Persons* (New York: Oxford, 1984), p. 96.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ARISTOTLE'S SOLUTION TO THE PROBLEM OF SENSIBLE SUBSTANCE*

THE classic problem with the central books of the *Metaphysics* is how to reconcile the apparently contrary requirements Aristotle places on substance. Of the four candidates for substance he lists in Z 3, essence is clearly the most successful. Essence is form, and many particulars apparently share the same form (1034a5–8, 1058b5–9). Yet, Aristotle also argues in Z 13–16 that nothing said universally can be substance. How can many things have the same essence if essence is not a universal? The problem is that essence seems to be both universal and individual.

Four approaches have been proposed: (1) Aristotle is inconsistent; (2) each composite individual has its own unique form, and so form is individual; (3) generic universals cannot be substances but species-forms can be, and so form is universal; and (4) form is neither.[1] The last, the position of Joseph Owens, is usually rejected as nonsensical. I shall advance another alternative: that form is both universal and individual, just as the text indicates, but that Aristotle uses these terms to refer not to kinds of entities but to characteristics. Form has both, because it is one in two different ways, and form can be one in both ways because it is actuality. So understood, Aristotle's account of form amounts to a kind of argument for identifying it as actuality.

I.

Let me begin with the requirement that essence or form be universal. Why should essence be universal? Early in his discussion of essence, at 1030a11–13, Aristotle declares that essence belongs to "nothing which is not a species of a genus." The reason is that only the species

* To be presented in an APA symposium on Aristotle's Metaphysics, December 30, 1987. Charlotte Witt will be co-symposiast, and Sarah W. Broadie will comment; see this JOURNAL, this issue, 673–679 and 679–681, respectively, for their contributions.

This is a condensed version of the final chapter of my *One and Many in Aristotle's Metaphysics: The Central Books* (Columbus: Ohio State UP, forthcoming). What I can only sketch here is expounded in detail there.

[1] Their defenders include: (1) J. H. Lesher, "Aristotle on Form, Substance, and Universals," *Phronesis*, XVI, 2 (1971): 169–178; (2) Rogers Albritton, "Forms of Particular Substances in Aristotle's *Metaphysics*," this JOURNAL, LIV, 22 (October 24, 1957): 700–704; (3) M. J. Woods, "Problems in Metaphysics Z, Chapter 13," in J. M. E. Moravcsik, ed., *Aristotle: A Collection of Critical Essays* (Notre Dame, Ind.: University Press, 1968), pp. 215–238; (4) Joseph Owens, *The Doctrine of Being in the Aristotelian 'Metaphysics'*, 3rd ed. (Toronto: Pontifical Institute of Mediaeval Studies, 1978), pp. 390–393. Many others should be mentioned, but there is insufficient space here.

0022-362X/87/8411/0666$00.70 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is "not said by something's being said of something else"; it is "primary" (1030a7–11). When Aristotle repeats his results later in the same discussion, he substitutes one for primary and substance for species (1030b4–10). Thus, the species is substance, because only what is one can have an essence and only species or substance is one. Species is one, because it has no parts and because the formula of what has no parts contains no additions. Thus, species is one through its formula; it is indivisible in formula (Δ 6, 1016a32–b6, b31–33). Not all species are one in formula and have essences. For Aristotle denies that accidents, accidental composites (like white man), *per se* attributes, and *per se* composites (like snub nose) have definitions and essences in the most proper sense. Their formulas contain additions. By default, only the formulas of species of substances are indivisible. To say that such a species *has* an essence is not to say that it *is* an essence. This latter follows, however, from the argument of Z 6, a section to which I shall return shortly. Apparently, substance or essence is a species.

There is also some support for universal forms in Z 10, though the text is a bit ambiguous. Aristotle denies that man and horse and other universals are substances (1035b27–30). They are composites of formula and "matter as universal." These composite universals—David Ross[2] calls them "materiate universals"—cannot be Aristotle's usual universals, because the passage goes on to speak of parts of form and parts of the composite of form and matter and to insist that there is a formula of only a part of the composite, the form or essence (1035b31–1036a1). Since he declares in virtually the same breath that the formula is of the universal, the form must be universal. It is not the composite's universality that prevents it from being substance but its plurality, its containing matter—just the theme of Z 10. Substance is a part of this composite, the universal form.

Still another reason to think that essence is universal is the account Aristotle gives of its unity in Z 12. At issue there is how that whose formula is a definition can be one (1037b11–12). The chapter shows that a formula can consist of a single constituent, the ultimate differentia, because this implicitly contains other differentiae and the genus. "The definition is some one formula and of a substance, so that it is the formula of something one; for substance is something one and a this" (1037b25–27). Since the formula is one, the substance it defines is one. But the formula is of the species; so, again, form must be universal.

[2] *Aristotle's Metaphysics*, I (New York: Oxford, 1958), p. c.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Finally, in Z 7–9 Aristotle describes the generation of a composite individual as the coming to be of a form in a matter, and he asks whether the form is generated in the generation of the composite. In order to avoid an infinite regress, some form and some matter must be ungenerated. The form pre-exists generation, because it exists in the agent that generates. As Aristotle puts it, "man generates man" (1033b32). As a result, any two instances of man are "the same, not in number but in species" (1033b31–32).

All these arguments show why essence cannot be individual or unique to an individual. First, if my form were unique, I could not have got it from my father: either form would be generated along with composite, or Aristotle would need to abandon his naturalistic account of generation. The human form we share is indivisible (1034a8); it cannot be further divided into individual species (cf. 999a4–5; 1058b5–10). Second, individual forms would be unknowable (cf. 1040a8–22), for we know when we grasp a formula and the formula is of the universal (1035b34–1036a1). Since Aristotle assumes that each essence has a formula, it cannot be individual.

Aristotle's distinction between "your form" and "my form" in Λ (1071a27–29) is often cited as evidence for individual forms. But what Aristotle claims there is that your form, matter, and mover differ from mine, though they are the same universally. That *all* these principles differ does not imply that they *each* differ. If "my form" is unique, it is because form has a secondary sense: like the "essence" of white man (1030b12–13), the secondary essence of Socrates is a composite of form, matter (taken universally), and attributes (1035b27–1036a8; 1058a10–11). Even if there are individual forms, there is no reason to think them prior to universal forms.

From all these arguments, it seems to follow that essence is universal. If we look closely at what Aristotle has really argued, however, what we find is only that essence must be one in formula. Species seems to be substance because, without material or formal parts, it is one in formula. Likewise, parent and child are one in species because they are one in formula (cf. 1016b31–33). Though Aristotle speaks of the species as substance, all he has shown is that whatever is substance must be one in formula.

II.

Let me now turn to the other requirement. Why should essence be individual? One passage that may seem to provide evidence for this is Aristotle's equation of thing (ἕκαστον) and essence in Z 6. If "thing" here were the composite (καθ' ἕκαστον), essence would be individual. But the equation would also be unintelligible, because the composite contains matter besides form. The "thing" equated with essence

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

cannot be the composite, because the equation holds only for things that are "*per se* and primary" (cf. 1031b13–14, 1030a7–11) and the composite lacks these characters. From the description of primary in Z 1031a29–b3, only substrates are primary. Since, as we know from Z 3, the material and composite substrates depend on form, only the formal substrate could be *per se* and one with essence. Thus, Z 6's identity of essence and thing is an identity of essence and form.

This doctrine explains a puzzling argument in Z 13 that nothing said universally is substance. Its key premise is, "things (ὧν) whose substance and essence are one are themselves one" (1038b14–15). The reason that this argument excludes universals is that, by definition, they are common to many things (1038b11–12). If a universal were the substance of these things, they would have one substance, and they should be one. Thus, no "one over many" can be substance. But this argument seems equally effective against Aristotle's essence: since many things share the same essence, their essence and substance would be one while they are many. This latter reasoning is mistaken, because the thing whose substance is essence is not the same as the thing whose substance would be the universal. The latter is the composite; the former must be that *per se* and primary thing we saw in Z 6, form, because form rather than the composite is what has an essence. Applied to form, the requirement of 1038b14–15 insists that when the essence is one, one thing, the form, be one. Form meets this requirement, because the requirement applies to it differently than it would have to apply to universals.

It is not surprising that Z 6 provides the key to understanding this argument in Z 13, for Z 6 supports its key premise. Because thing and essence are one, unity of essence implies unity of thing. Aristotle's statement of the requirement later in a similar argument (1040b17) shows that unity here must be numeric, a point that we could infer from Z 13 anyway. This is very odd, to say the least, for it means that the universal needs to be one in number before the requirement could exclude it. How could the universal be one in number? Aristotle's point is not that universals are numerically one but that they would need to be numerically one to be substances. Were a universal, such as being, the substance of its instances, many things would have (numerically) one substance, being. If it were a substance, being would be an entity that is numerically one, but this would lead to contradiction (1040b16–19). If this interpretation is correct, Aristotle's argument against universals excludes all of them, even the species (*pace* Woods and Albritton). In contrast, essence is able to be numerically one without contradiction. The argument assumes that substance must be numerically one or individual.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Most of Aristotle's other arguments against universals turn on two characters: universals are predicated of substrates and they are composed of other universals. If universals were substances, substances would be composed of other substances. What would be wrong with this? Aristotle explains that, were the universal to have a plurality of actual constituents, it would not be one in number (1039a4–6). Again, substance is one in number; it is a self-subsistent individual.

From all of this, it follows that no universal is substance (1042a21–22; 1053b16–21). Substance must be some sort of individual, but what sort? What Aristotle has argued is that form must be one in number. Initially, the universal seems to be substance most of all (1038b2–7) because it is one, but it is not one in the pertinent way. Assuming that it is numerically one, as the Platonists do and as we must if it is to be substance, generates false consequences. Universals lack the numeric unity required of substances.

Let me now summarize. I have directed your attention to Aristotle's arguments that form is not individual but universal, and also to his arguments that form is not universal but individual. What we saw as we looked closely was that Aristotle really argues not about universality or individuality; he argues that form is one in number and in formula. Elsewhere he equates individual and one in number and suggests the parallel equation between one in formula and universal (B 4, 999b33–1000a1). So the result of Z–H is that form or essence is *both* universal and individual. In the usual senses of these terms, this claim would be nonsense, but here they refer not to kinds but to two ways of being one. Since these ones are not contrary (1016b36), the demands that the central books place upon substance need not be contrary.

III.

The preceding discussion removes the difficulty without solving the real problem. Since something that has one matter is both numerically one and one in formula (1016b36), the unity demands need not be contrary. But essence does not have one matter. If it can meet the demands, it must do so in another way. How can essence have both types of unity? Just what sort of thing would it need to be? Aristotle's ascription of seemingly contrary characteristics to form reflects neither confusion nor shift in doctrine; it is standard aporetic method. It supports the identification of form as actuality and matter as potentiality, for these identifications allow form to be one in both ways. Aristotle's extensive criticisms of Platonic form, including his arguments that that form cannot be one in both ways (1039a24–b6; 1002b22–25), help to draw up the *aporia*. He identifies form as actuality in book H to explain how the composite can be one; this and

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the treatment of actuality in book Θ show how form can be one in both ways.

The term "actuality" (ἐνέργεια, ἐντελέχεια) refers to the function that makes an entity what it is. Perfect examples are the pure actualities, the unmoved movers. These are entities whose being consists entirely of their function (1050b16–17; 1071b19–22). They are complete; their ends lie within themselves. Aristotle contrasts them with motions, processes whose ends lie outside of themselves (1048b18–35). Sensible substances occupy an intermediate position. They are forms and actualities, but forms that cannot exist apart from matter and forms that come to be present in and pass out of matter. Although Aristotle calls them actualities, they retain a close connection with matter. He also refers to them as motions (1026a1–2; 1036b28–30), and he suggests that they are potentialities (1050b11–13). The standard definition of man in the *Metaphysics,* two-footed animal, provides a way to understand these actualities. Having two feet is both the actuality that makes us what we are and a capacity for further activity, walking in a characteristically human way. It is what the *De Anima* terms a "first actuality" (412a21–b1). Although it is not, in Aristotle's more considered view, our function, I think that the *Metaphysics* chooses it as a standard example because it illustrates the identity of form and proximate matter in a way that the more accurate human actuality, capacity for thinking, may not.

Form's inability to exist apart from matter has a variety of consequences. First, the form must come to be present in a matter. So the processes of birth, growth, and development are natural motions whose end is the form. Second, because matter inevitably wears down, decay and death are also natural motions. Consequently, any form of life must reproduce to sustain itself. That is to say, form inevitably individuates itself in matter. Individuation is a natural process for any form of life. I have chosen Wittgenstein's phrase "form of life" intentionally, because it captures the particularity of form rather than its collective nature. We can look at the processes of birth, growth, reproduction, and death from the perspective of the form; these are all inevitable aspects of form's manifestation of itself. Given form's nature, these processes must occur: form must have a plurality of instances and, through this plurality, form attains a completeness that imitates the completeness of pure actualities. I am reminded of E. O. Wilson's remark that an individual person is nothing more than a vehicle for genes. Genes and forms differ in crucial respects, but the argument of the central books of the *Metaphysics* forces us to make a parallel gestalt shift away from the pri-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

macy of individual composites to the perspective of form. Once we grasp the form as an activity that necessarily occurs in matter, the particularity of individual matters becomes irrelevant. Form's presence in matter is simply its peculiar mode of being. From the perspective of form, individuals are merely vehicles for its continued existence.

With this we can see how form is both one in number and one in species. Aristotle argues that the first unmoved mover is one in number, because it is an actuality that lacks matter; it is apparently one in formula for the same reason (1074a31–37). The actualities that he identifies as forms of sensibles are also numerically one, not simply one, but one to the extent to which they are separate from matter (1042a28–29). Because these actualities contain no material or formal parts, they are also indivisible in formula. Although forms of sensibles lack the degree of unity possessed by pure actualities, they do meet the unity requirements for substance. The central books use these requirements to determine the nature form must have, and what emerges is the character of a peculiar entity, actuality. This form is individual without being in a single place or time; it is universal without being predicable of individuals (cf. 1049a34–b2). It is one nature inevitably manifested in a plurality. Because only such an entity *can* satisfy the unity requirements of substance, it *must* exist. Since this actuality is neither a predicate nor a subject, as these are usually conceived, it cannot be known through demonstration. Even our intuition of it is obscured by its presence in an accidental and thus unknowable plurality. Form's partial unintelligibility stems from its nature rather than our deficiencies, a point Aristotle may well have in mind when he declares that the question of substance will always remain puzzling (1028b2–4).

EDWARD HALPER

University of Georgia

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

HYLOMORPHISM IN ARISTOTLE*

WHAT is the relationship between form and matter in an Aristotelian composite substance? The interpretation of hylomorphism I develop below is based on two ideas. First, matter and form should be understood in terms of potentiality and actuality; second, potentiality and actuality should be understood adverbially, as ways of being. Very briefly, my interpretation holds that matter is one way of being the composite substance, and form is another way of being it. This understanding of matter and form solves the problem of the unity of the composite substance, because actuality determines potentiality for Aristotle. The bulk of this paper is devoted to spelling out what might be meant by "ways of being," and how it is that actuality "determines" potentiality.

In Aristotle's discussion of being in *Metaphysics* v, 7, the distinction between actual and potential being is presented last. Aristotle tells us that the distinction applies to the other three divisions of being: essential being, accidental being, and being as true and false (1017a35). To see what Aristotle might have in mind, let us consider essential being. Essential being is mapped out by categorical predications, e.g., "Socrates is a human being." If we apply the distinction to the categorical predication, we get: "Socrates is potentially a human being," and "Socrates is actually a human being." But what does it mean to say that something is potentially a human being? Is it a human being or not? And what does it mean to say that something is actually a human being? How does that assertion differ from simply saying that something *is* a human being?

In the brief text we are considering, Aristotle illustrates the potentiality/actuality distinction as follows:

> For we say both of that which sees potentially and of that which sees actually, that it is seeing, and both of that which can use knowledge and of that which is using it, that it knows, and both of that to which rest is already present and of that which can rest, that it rests. And similarly in the case of substances we say that Hermes is in the stone and the half of the line is in the line, and we say of that which is not yet ripe that it is corn (1017b2–9).

* To be presented in an APA symposium on Aristotle's Metaphysics, December 30, 1987. Edward Halper will be co-symposiast, and Sarah W. Broadie will comment. See this JOURNAL, this issue, 666–672 and 679–681, respectively, for their contributions.

0022-362X/87/8411/0673$00.70 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The first set of contrasts is between a capacity and its exercise. A person who has knowledge has a certain capacity. In relation to the exercise of that knowledge, the person is a potential knower. Here, the person who is an actual knower is a person exercising that capacity. Let us group together these sorts of examples of potentiality and actuality under the label *capacity and exercise* (CE for short). In relation to these examples, what we mean when we say "Socrates is actually a human being" is that he is exercising his capacity for being human. He is in action.

Aristotle next lists several examples of substances; unripe corn is potentially corn, when it ripens it is actually corn. This example invites us to think of, not a capacity and its exercise, but a continuum of directed development. Unripe corn is corn potentially—it is in process toward its end which is being corn actually. Mature corn is corn actually—it is in its end. If I were a German philosopher, I would call this application of potentiality and actuality *being-toward-its-end* and *being-in-its-end*. Instead, I group together these examples of potentiality and actuality under the label *process and end* (PE for short). In relation to these examples, what we mean when we say "Socrates is actually a human being" is that Socrates has fully achieved that set of capacities which a mature human being has. Of the youthful Socrates, the Socrates who lacks the distinctively human capacity to philosophize, we would say "Socrates is potentially a human being."

It should be noted that, in our text, Aristotle stresses the similarity of the PE examples to the CE examples rather than their differences (1017b6). How is the relationship between unripe corn and mature corn like the relationship between a capacity and its exercise? Here we must distinguish two capacities: the capacity for becoming corn and the capacity for being corn. The capacity for becoming corn is the capacity the seedling has to grow, develop, and ripen. The capacity to develop into ripe corn is one which the seedling has only so long as it is potentially, and not actually, ripe. The potentiality for becoming corn is used up in the corn's development. Hence, the potentiality for becoming corn differs from the relationship between a capacity and its exercise. The potentiality for becoming corn perishes as the seedling ripens, but a person retains the capacity for sight even while she is seeing.

This observation suggests that Aristotle is thinking rather about the seedling's capacity for being corn. The seedling's capacity for being ripe corn is a potentiality the corn has to achieve an end. This capacity is exercised when the corn is ripe and mature, just as a person's capacity for sight is exercised when she sees. The cases are

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

similar, because in neither is the capacity used up in the exercise. These points suggest that potentiality and actuality in connection with being are not used in an analysis of the changes or motions that entities undergo. It is not the motion of ripening which Aristotle is trying to explain by contrasting unripe and ripe corn, and calling one "potential" and the other "actual," but rather different ways of being corn. This is an important point, because potentiality and actuality do figure centrally in Aristotle's analysis of motion, and so it would be easy to misunderstand his point.

If the ripening corn example of potentiality and actuality is not understood in terms of an analysis of the motion of ripening, then another set of questions arise concerning the meaning of the distinction. What does Aristotle mean by distinguishing different ways of being corn or being human? Why does he make this distinction? Why not just stick to austere categorical being, and say of the unripe corn and the child what we would say, namely that it is corn or a human being?

The answer to this question is that Aristotle thinks of these sorts of beings—composite substances—as being *en route* to their respective ends. Beings that are *en route* are just beings at differing levels of potentiality or actuality. Here again it is important not to take his point as being about the route, the series of motions which accomplish the perfection of the corn or the human. Rather, with the notions of potentiality and actuality, Aristotle expresses ontological relations among the beings at different stages, from a perspective in which the stages are all directed toward a series of ends. The phrase 'directed toward a series of ends' is important. For, viewing the life of a human being as comprised of a succession of stages, youth, middle age, old age, does not have any obvious ontological implications. If all you mean by saying that a youth is potentially middle-aged is that the former stage preceeds the latter, then a young person differs from an older person only by temporal position. On this view, the earlier stage is fully what it is quite independently of the later stage. For Aristotle, however, there is a connection between the stages, because the *way* the entity is what it is in the earlier stage is determined by a later stage, an actuality. The stages are determined by the overall end of the progression. So, the youthful Socrates is potentially human, where what it is to be human is given in the final actuality, the activity of a mature human being.

I said that the being of the entity in the earlier stage is *determined by* a correlated end. What sort of determination is in question here? In *Metaphysics* IX, 9, Aristotle tells us that actuality is prior to potentiality in several ways. Two priorities, priority in being and in defini-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tion, are what I have in mind in saying that potential being is determined by actual being.

Concerning the priority in being of actuality in relation to potentiality, Aristotle says that "everything that comes to be moves towards a principle and end. For that for the sake of which a thing is is its principle, and the becoming is for the sake of the end" (1050a7–9). Beings that are *en route,* exist toward their ends. They are for the sake of their ends. The principle of being for the youthful Socrates is an actuality, namely Socrates actively being human. But, in the young Socrates, the principle that governs his being, his form or end, is not actually or fully present. So, the youthful Socrates is potentially human, precisely because his being is determined by an end or actuality that he has not yet achieved. Being potentially human is one way of being human, a way in which what you are is not yet realized; your principle or end is not yet realized in you. Being actually human is another way of being human, a way in which what you are, the principle of your being, is fully realized in you.

Actuality is also prior to potentiality in definition, because a potentiality is defined in relation to its correlative actuality. For example, having the capacity of sight is defined as being able to see, to engage in a certain activity (1049b13–15). Similarly, when we say what the youthful Socrates is, we say he is potentially human, he is potentially a being that engages in a certain characteristic range of activities. The priority of actuality over potentiality in definition is really a consequence of the priority in being. Since an Aristotelian definition says what a thing is and since the principle of being of a potentiality is the corresponding actuality, it follows that the statement of what the potentiality is, its definition, will refer to the end or actuality.

Potential and actual being describe the ways in which things are what they are. In Aristotle's system, they are a necessary addition to the more familiar picture of categorical being, since they describe a basic feature of reality, namely that it is directional. What something is is determined by its form or end, and its relationship to that form or end. Potentiality and actuality admit of degrees, because an end can be realized in different ways, i.e., to a greater or lesser extent. The way of being of an entity, namely the degree to which it has realized its form or end, has two varieties. The mature Socrates is fully in his form and the immature Socrates is not; Socrates in action is more fully realizing his form than Socrates asleep.

In my Socrates example, I have been thinking of a human being as existing in different ways, at various degrees of actuality. Human beings, composite substances, have this sort of being, because they are composites of matter and form. "Further matter exists in a po-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tential state, just because it may attain to its form; and when it exists actually then it is in its form" (1050a15–16). Underlying the process of the composite substance realizing its form more and more, is matter, which at one moment "may attain to its form" and at another moment is "in its form." Even when matter is in its form, however, the composite substance that results has two aspects, two ways of being. One is potential being associated with matter, and the other actual being associated with form. If it did not have these two aspects, then it would not be a being capable of degrees of actualization. But how ought we to think of this fundamental duality of aspect in the composite? And how does it mesh with the other applications of potentiality and actuality? And what kind of unity results?

What we are looking for is a level that underlies PE and CE potentiality and actuality. For Aristotle, a human being's form is his or her soul, and in the *De Anima* soul is defined as "the first actuality of a natural body that has life potentially" (412a20–21). The matter of an ensouled entity is "a natural body that has life potentially"; the matter of a human being is a natural body that has human life potentially. Form or soul is the first grade of actuality which corresponds to the capacity for life, in our example, for human life. A composite substance like a human being has two ways of being; its matter is a body that is potentially alive and its form is a soul or a set of capacities that constitute human life. Now these capacities are developed throughout life, from the immature specimen of a human being to a mature exemplar. This development, what I called PE potentiality and actuality, is possible, because one aspect of the composite is material, potential, has a potentiality for those capacities whose principle is soul or form. If the matter did not have the right set of potentialities, then the process of maturity or actualization would not occur. On the other hand, if the composite did not retain its potentiality, its being as matter, then there would be no *process* of maturity, no gaps in activity, no eventual death, and no problem of unity (*Metaphysics* VIII, 6 1045b23–24).

In living out his or her life a human being is sometimes active and sometimes at rest. Human life is a set of capacities that we exercise intermittently. Life and living are an example of a capacity and its exercise. And, according to Aristotle, soul is the source or origin of these human capacities. It is that aspect of the composite which causes it to be alive, to have that set of capacities. The matter, as I have said, is only potentially alive. It is the capacity for the capacity to live. One might then think of matter and form in the composite on the model of a capacity and its exercise, where the exercise is just the actual having of the capacity. The difference between the two cases is

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that the matter itself is never actually alive; its potentiality for life is retained. The human being, the composite, is alive and intermittently active. Its material being is a drag on the system, and is not itself actualized by being in its form.

In the *De Anima,* Aristotle illustrates this level of potentiality, the capacity for a capacity, with the example of a knower:

> But we must make distinctions concerning potentiality and actuality; for at the moment we are speaking of them in an unqualified way. We can speak of something as a knower either as when we say that man is a knower meaning that man falls within the class of beings that know or have knowledge, or as when we are speaking of a man who possesses a knowledge of grammar. (Each of these is potential, but not in the same way, the one being a potential knower, because his kind or matter is such and such, the other because when he wants to he can contemplate, if nothing external prevents him.) (417a22–28)

The potential knower in the first sense has the capacity for a capacity. A human being, because he or she *is* a human being, has a capacity to have knowledge of various sorts. For Aristotle, the having of knowledge of various sorts is also a capacity, or potentiality, in relation to its exercise. It is an example of CE potentiality and actuality. A knower has the potentiality to have knowledge, Aristotle tells us, because "his kind or matter is of a certain sort." So the capacity for the capacity that is having, but not exercising, knowledge is retained by the knower so long as he or she is a human being. In the case I am interested in, the sort of potentiality associated with matter in a composite substance, a capacity for the capacity for actively living a human life, is retained so long as the matter is of the appropriate sort, i.e., potentially alive and human.

As in the other examples of actuality and potentiality, in the case of matter and form, actuality determines potentiality. As before, the determination is of two sorts—actuality is prior in being and in definition to potentiality. Actuality or form is prior to potentiality or matter in being, because the matter exists toward the end or form. In discussing the priority in being of the actual over the potential, Aristotle says that "matter exists in a potential state, just because it may attain to its form; and when it exists actually, then it is in its form" (*Metaphysics* IX, 8 1050a15–16, also 1088b1, 1092a3–5).

Furthermore, the matter of a composite substance is defined, in so far as it has a definition, by reference to a corresponding actuality. Why can we not define matter quite independently, in terms of its own material qualities? The answer to this question is that Aristotle

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

thinks that a definition should tell you what a thing is, and matter exists or is toward or for the sake of form. And so it should be defined as potentially enformed. In talking about various approaches to definition, Aristotle says: "And so, in defining, those who define a house as stones, bricks and timbers are speaking of the potential house, for these are the matter" (1043a14–16). To view the matter as the potential house is a consequence of how Aristotle sees reality ordered; it is directional, which means in this instance that matter, potentiality as matter, is toward form, just as the substances themselves are directional, they exist toward their forms.

Aristotle thinks that, viewed as potentiality and actuality, form and matter are one. Moreover, he states this conviction in a context in which he is discussing the problem of the unity of the object of definition. To see that matter and form are to one another as potentiality is to actuality, is to see that they are neither two definitionally independent objects nor two definitionally independent parts of a whole. It is because what a potentiality is is dependent upon what an actuality is, that composite substances are unified objects of definition. Matter is definitionally dependent upon form. There is, therefore, only one definition; matter and form as ways of being of the substance satisfy that definition to different degrees.

According to my interpretation, Aristotle's hylomorphism is an application of his distinction between potentiality and actuality to composite substances, and so the intelligibility and plausibility of hylomorphism hinges upon whether or not it is plausible to think that there are ways of being in addition to kinds of being.

CHARLOTTE WITT

University of New Hampshire

PROBLEMS OF ARISTOTLE'S CONCEPT OF FORM*

Edward Halper and Charlotte Witt are concerned with the same general topic, Aristotle's conception of form or essence in the central books of the *Metaphysics*. Both scholars seek to make sense of this in terms drawn from Aristotle himself—no doubt on the principle that an external assessment should be based on sound exegesis. Despite the common focus, their papers neither overlap nor conflict, since they deal with different difficulties. I shall summarize each and raise a few questions.

* Abstract of a paper to be presented in an APA symposium on Aristotle's Metaphysics, December 30, 1987, commenting on papers by Edward Halper and Charlotte Witt, this JOURNAL, this issue, 666–672 and 673–679, respectively.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Halper's is the classic problem of how form or essence can be, as the texts indicate, both "individual and universal." Halper refuses to underplay the evidence for either side of the dilemma, but also refuses to accept the two sides in the terms in which they are usually cast. Aristotle's doctrine, he argues, is not primarily that form is both universal and individual, but that it is unitary and a source of unity in just the ways that qualify it to count as substance. (It helps if we bear in mind that Aristotle's claim that form is substance amounts, in this context, to the claim that form is the principle of substantiality of composite individuals.) Halper argues (a) that the passages implying that form is universal are meant to show that form is one *in formula,* and (b) that those which deny it to be a universal mainly make the point that a universal cannot ground the *numerical* unity of a composite individual. Thus, the same quest for unity motivates both sides of the dilemma: a fact which highlights rather than reduces the paradox. Halper goes on to sketch (c) a resolution in terms of the concept of actuality. Forms as the actualities of composites like Callias and Bucephalus must particularize themselves in matter. Thus, a form as such is not at any one particular place at a time, hence (ci) is a universal. On the other hand, Halper claims, form is also (cii) not a universal, because it is not predicable of composite individuals.

The arguments for (a) seem to identify the substance of a composite with its species, hence with a universal. Is each actuality [as per (c)] a species? If not, what is the relation? What is the criterion for being a universal? Is it not being at one place at a time, or is it being predicable of a composite? As to the last point, Aristotle surely holds that the form or essence *is* predicable of the sensible individual, as well as of its matter (thus the definition of *man* is predicable of Callias). Perhaps rather than (cii) Halper means that it is not *qua* thus predicable that form functions as the substance of the composite.

Witt's problem is the unity of form and matter in the composite. She argues (a) that matter and form should be understood as potentiality and actuality, and (b) that potentiality and actuality are "different ways of being." Potentiality is to actuality either as capacity to exercise (CE) or as potentiality to end (PE). In PE the potentiality (which is a potentiality for coming to be "in" the end) is used up in the achievement of the actuality; in CE this is not so, and the capacity may even be enhanced by the exercise. In both cases actuality is prior in being and definition. The matter of a composite substance stands to the form as capacity to exercise, in Witt's view, except that in this case the "exercise," or form, is itself a set of capacities for active

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

functioning. Matter, therefore, as potentiality, is not definable without reference to form, nor is it ontologically independent. Thus, the unity of form and matter is not the insoluble problem it would be if matter and form were each something in their own right.

My question concerns the sense of the potential/actual contrast as applied to matter and form. Presumably the PE sense is unsuitable because the potential here is superseded by the actual, whereas the problem of unity only arises because matter and form are assumed to be present together in the composite. But the CE sense seems in conflict with *De Anima,* which says that body (matter) stands to soul (form) as the potentiality of knowledge in an ignoramus to the possession (whether exercised or not) of knowledge learnt. But that is a case of PE. In this an inessential feature of the example? If not, Witt's suggestion that matter stands in a CE relation to form is a departure from Aristotle, although a reasonable one, given that one wants to say that the matter of the composite is not done away with by the presence of the form. This certainly seems a strange thing not to want to say, when one thinks of the matter as an organism's body and the form as its soul. But in some passages of the *Metaphysics* Aristotle seems to deny that matter is present as a thing at all (even a potential thing) in the composite, for he gives it the status of an adjective (the bed is wooden, not of wood, etc.). By this move he does not so much solve the problem of unity as make sure that it does not arise.

SARAH BROADIE

Yale University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE DISCOURSE OF MODERNITY: HEGEL AND HABERMAS*

The status of the "modern project" is strongly contested today; under such summary labels as "modernism versus postmodernism" or "enlightenment versus deconstruction," champions and critics of the project are embroiled in lively skirmishes both inside and outside of academia. Unfortunately, the salience of the issues is not often reflected in the character of exchanges: more than other contemporary topics, the themes of modernity and postmodernity tend to be submerged in trendiness and facile rhetoric. Only rarely does literature devoted to these themes rise above the plane of broad manifestoes or petty polemics; Jürgen Habermas's *The Philosophical Discourse of Modernity*[1] is clearly such an exception. First published in 1985, the study lends substance and focus to otherwise often diffuse debates; moving beyond surface changes and technical gadgets, it locates the core issues of modernity on the level of philosophy—more specifically of an ongoing "philosophical discourse." From Habermas's perspective, modernity is intimately linked with the central aspirations of the Reformation and Enlightenment: the aspirations of cognitive rationality, moral autonomy, and social-political self-determination. Accentuating this linkage means to underscore the stakes involved in current discussions. For, profiled against the backdrop of heteronomy and caprice, how can one blandly dismiss modernity's gains? On the other hand, given the lengthening shadows cast by anthropocentrism and technical prowess, how can one blithely endorse these gains without naiveté or callousness?

One of the chief merits of Habermas's study is the treatment of modernity not as a platform or doctrine but as a discourse or conversation—a conversation made up of different protagonists or voices and stretching over successive historical periods. In Habermas's presentation, the discourse was inaugurated by Enlightenment thinkers from Descartes to Voltaire and first crystallized in the rationalist

* To be presented in an APA symposium on Enlightenment and Rationality, December 30, 1987. Axel Honneth will be co-symposiast, and Tom Rockmore will comment; see this JOURNAL, this issue, 692–699 and 699–701, respectively, for their contributions.

[1] *Der philosophische Diskurs der Moderne* (Frankfurt-Main: Suhrkamp, 1985). All the translations are mine. Cf. *The Philosophical Discourse of Modernity: Twelve Lectures*, Frederick G. Lawrence, trans. (Cambridge, Mass.: MIT Press, 1987).

0022-362X/87/8411/0682$01.10 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

theories of Kant and Fichte. In Kantian thought, he notes, modernity meant basically the progressive refinement of consciousness and subjectivity, specifically the segregation of reason into the domains of science, ethical freedom, and aesthetic judgment—a segregation apparently achieved without costs or charring effects. Neglect of these costs soon led to dissent or insurgency within modern discourse, in the sense that the *cantus firmus* of analytical rationalism was joined by the supplementary or "counter discourse" of rational synthesis —an insurgency which found its chief voice in Hegel's system. Countering the divisions and cleavages (*Entzweiungen*) resulting from modernity, Hegel—without abandoning the modern project —sought to reconcile the dichotomies of "nature and spirit, sensuality and reason, *Verstand* and *Vernunft,* theoretical and practical reason, judgment and imagination, finitude and infinity, knowledge and faith" (32). While praising Hegel's philosophical élan, Habermas finds Hegel's insurgency flawed and ultimately unsuccessful—mainly because of its "subjectivist" moorings and its excessively theoretical contemplative character. During the nineteenth century, the Hegelian legacy was continued by the opposing camps of "Young Hegelians" and "Right Hegelians": the first devoted to the implementation of reason on the basis of praxis or productivity, and the second to the maintenance of objective rational structures as embedded in state, economy, or technology. As opposed to these internal modulations of modernity, the study shifts attention at this point to another, more radical insurgency or antidiscourse (*Sonderdiskurs*)—one which seeks not so much to modify as to "cancel" the modern project and whose diverse articulations take up the bulk of the volume. The instigator of this radical stance was Nietzsche who, in the study, figures as the turning point or as the "turntable" (*Drehscheibe*) ushering in the move from modernity to postmodernism. As in Hegel's case, Nietzsche's legacy is said to have been continued by two opposing (but also interdependent) camps: one pursuing a more skeptical and quasiscientific approach, the other attempting an ontological or quasi-ontological reversal of modernity. The chief representatives of the first camp are Georges Bataille, Jacques Lacan, and Michel Foucault; those of the second camp Martin Heidegger and Jacques Derrida (120).

To be sure, the preceding synopsis does not do full justice to the complexity of the study, particularly to its many sidelines and repeated interludes or "excursions"; nevertheless, it captures, I believe, the chief strands and counter strands of the modern discourse (and antidiscourse) as seen by Habermas. It cannot be my ambition in the present context to recapitulate and discuss all the facets of this

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

discursive or conversational fabric; such an effort would require the format of a full-length commentary or counter study. In the following pages, I propose to concentrate on one central episode in the study's historical scenario, the episode connected with the name of Hegel or Hegelian philosophy.[2] The choice of this episode seems justified by its crucial role in Habermas's narrative: Hegel inaugurated the modern discourse in its multidimensionality and thus established the parameters for the continuation of this discourse by subsequent thinkers and schools of thought. By erecting these parameters, Hegel also serves as the basic foil or backdrop against which the later antidiscourse, initiated by Nietzsche and pursued by his heirs, is prominently profiled.[3]

I.

The Renaissance and Reformation (together with the discovery of the "New World") heralded an implicit break with the classical and medieval past; but the notion of a distinctly "modern" period emerged only slowly in the aftermath of these events. According to Habermas, it was left to Hegel to grasp the philosophical meaning and import of the modern project. Hegel, he writes, was "the first philosopher to develop a clear conception of modernity" (13). Although anticipated dimly by Enlightenment thinkers from Descartes to Kant, it was only toward the end of the 18th century that "the problem of the *self-understanding* of modernity became so acute that Hegel could perceive it *as* a philosophical problem and moreover as *the basic problem* of his philosophy" (26). Together with his philosophical precursors, Hegel located the core of modernity in the principle of "subjectivity"—a principle which carried for him mainly the connotations of individualism, critical-rational competence, and autonomy of action. The same principle had already been succinctly pinpointed by Kant, who treated subjectivity as the foundation of the segregated modern domains of science, (categorical) ethics, and expressive art. Kant, we read, "put in place of the substantive rationality bequeathed by metaphysics the notion of a reason differentiated into separate moments whose unity has merely formal character; he distinguished the faculties of practical reason and judgment from theoretical cognition and assigned to each moment its own place"

[2] In the APA symposium, I shall also discuss Nietzsche, Heidegger, and, in greater detail, Habermas's model of communicative rationality.

[3] I have discussed some of the other thinkers included in the study elsewhere (in a manner deviating from Habermas's exegesis): compare my "Pluralism Old and New: Foucault on Power," in *Polis and Praxis* (Cambridge, Mass.: MIT Press, 1984), pp. 77–103; and "Hermeneutics and Deconstruction: Gadamer and Derrida in Dialogue," in *Critical Encounters* (Notre Dame, Ind.: University Press, 1987), pp. 130–158.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(29). Yet, while pursuing his analytical task, Kant did not grasp the differentiation of reason as a problem, nor the separation of modern "value spheres" as division or a source of divisiveness (*Entzweiung*); consequently, he also ignored the synthetic "need" emerging from his analysis. Here was precisely the *motif* of Hegel's insurgency: while accepting the principle of subjectivity, he recognized both its emancipatory potential and its ambivalence. In Habermas's words, the principle "explained for him simultaneously the superiority of the modern world and its crisis character, in the sense that it represents both a world of progress and of alienated spirit. For this reason the first attempt to conceptualize the modern era was at the same time a critique of modernity" (27).

In the chapter devoted to Hegel, Habermas traces the successive stages of his insurgency within the confines of Enlightenment discourse. As he notes, the initial impulses of his synthetic efforts can be traced back to critical or "crisis experiences" of the young Hegel himself, experiences which nurtured his conviction "that reason must be marshalled as conciliatory power against the divided positivities of his age" (33). Turning first to Hegel's early (especially his theological) writings, Habermas points to a certain Romantic or "mythopoetic" version of reconciliation which Hegel shared with Schelling and Hölderlin, his friends in the Tübingen seminary. Countering both the orthodoxy of "positive" (or established) religion and the abstractness of Enlightenment ideas, these writings appealed to a purified public faith or civil religiosity as the bond tying together and reconciling the conflicting segments of society. Only when represented in public festivals and cults and linked with myths engaging heart and phantasy—Hegel argued at the time—could a religiously mediated reason "permeate the entire fabric of the state." The same writings also spoke of a "nexus of guilt" or a "causality of destiny" as the driving motor propelling a reconciliation of criminally severed relationships, a motor revealing the injury inflicted on others ultimately as self-injury. According to the study, Hegel "opposed to the abstract laws of morality the very different rule mechanism of a concrete nexus of guilt generated by the sundering of a prior ethical totality (*sittliche Totalität*)" (41); by suffering the consequences of his action, the criminal comes to recognize in the injured alien existence "his own repudiated nature." Shifting from narrative to critique, Habermas at this point challenges the character of the invoked totality or social bond: both in the case of civil religion and of the nexus of guilt, he argues, reconciliation relies on premodern life-forms which the process of modernization necessarily leaves behind. "For the fated reconciliation of a divided mo-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

dernity," he writes, "Hegel presupposes an ethical totality which is not germane to modern conditions but rather is *borrowed* from the idealized past of early Christian communities and the Greek *polis*"; yet, modernity had gained its self-understanding precisely through "a reflection which bars the systematic regress to such exemplary traditions" (42).[4]

Similar or related dilemmas—dilemmas soon recognized by Hegel himself—beset another early or transitional work in his intellectual development: the so-called "oldest system program" formulated in Frankfurt and still under the influence of Schelling and Hölderlin. In that program, the function of reconciliation was attributed to art or artistic-poetic imagination. "Rational religion," the study comments, "was presumed to yield to *art* in order to develop into a popular religion; the monotheism of reason and heart was to ally itself with the polytheism of imagination to produce an (aesthetic) mythology of ideas" (44). As Habermas observes, this program was clearly reminiscent of Schiller's letters (1795) on the "aesthetic education of mankind," and it was retained in Schelling's transcendental idealism and in Hölderlin's works to the end. According to Habermas's narrative, however, Hegel quickly abandoned this outlook as an insufficient remedy for modern ills and divisions. Given that modernity is based on subjectivity and critical reflection, only philosophical reason or thought—a thought moving within subjective reflection and beyond it—could accomplish the hoped for reconciliation and subdue the pitfalls of a solipsistic or domineering subjectivism. This insight, Habermas argues, was the crucial stepping stone to Hegel's notion of *absolute spirit*. If modern advances are to be taken seriously and yet to be corrected, he writes, "reason must indeed be construed as self-relation of a subject, but now as a reflection which does not simply impose itself on otherness through the pure force of subjectivity but rather as one which has its essence and motor only in the effort to oppose all finite absolutisms and to overcome all encountered positivities" (46). In contrast to static metaphysical conceptions of the past, the absolute spirit in Hegel's treatment consisted "purely in the process of the relation of finitude and infinity and thus in the consuming activity of self-discovery"; moving beyond the level of substances and fragmented subjects, the absolute is construed "solely as the mediating process of an unconditionally self-productive self-relation" (46). In this manner, Ha-

[4] As Habermas adds: "No matter how forcefully interpreted, the ethos of *polis* and early Christianity can no longer furnish the standard which could guide an internally divided modernity" (43).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

bermas concludes, Hegel utilized the philosophy of subjectivity "with the aim of overcoming a subject-centered reason; with this move the mature Hegel can criticize the defects of modernity without appealing to any *other* premise than the immanent-modern principle of subjectivity" (46).

The self-transcendence of modernity encapsulated in the absolute spirit is replicated, in somewhat different guise, on the level of *objective spirit,* and especially in the theory of the modern state. Turning to the *Philosophy of Right* (and earlier preparatory writings), Habermas finds Hegel's main contribution in his formulation of the notion of *civil society* as a domain differentiated from, and mediating between, family and state; 'civil society' in this context meant a mode of association governed by private interest and market exchanges. In formulating this notion and juxtaposing it to the state, Hegel took account both of the advances of modernity and its divisive effects. In Habermas's words, the issue confronting Hegel was "how civil society could be conceived not only as a *sphere of decay* of substantive ethics but in its negativity at the same time as a *necessary moment* of ethics (*Sittlichkeit*). He took his departure from the premise that the classical ideal of the *polis* cannot be restored in the context of modern, depoliticized social life; on the other hand, he maintained the idea of an ethical totality which he had thematized earlier under the rubric of popular religion" (51). By separating and simultaneously linking society and state (embodying the objective spirit), Hegel's *Philosophy of Right* promoted a self-transcendence of modernity under modern auspices; in opposing both a homogeneous *polis* and an unlimited sway of private interests, the work set itself apart "from restorative philosophies of the state as well as from rational natural law" (in the Enlightenment sense). To quote Hegel himself: "The idea of the modern state has this immense strength and depth that it allows the principle of subjectivity to unfold to the extreme of self-sustained individual separateness while simultaneously guiding the principle back into substantive unity and maintaining the latter within itself."[5]

II.

Having thus outlined Hegel's "mature" position as reflected in the notions of absolute and objective spirit, Habermas immediately proceeds to criticize this position for failing to solve modern predica-

[5] *Philosophy of Right,* T. M. Knox, trans. (New York: Oxford, 1967), p. 161 (par. 260); translation slightly altered.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ments. Two (related) reasons are given for this failure. Although having previously applauded Hegel's firm adherence to the modern "principle," Habermas now takes him to task for remaining hostage to a self-enclosed subjectivity unable to perform a synthetic function. "With the notion of the absolute (spirit)," we read, "Hegel regresses behind the intuitions of his youth: he conceives the overcoming of subjectivity only within the limits of the philosophy of the subject" (33).[6] By claiming the power of synthesis for absolute reason or subjectivity, Hegel is said to have feigned reconciliation by a slight-of-hand: "He would have had to demonstrate, rather than merely presuppose that (absolute) reason—which is more than abstract *Verstand—can* strictly reconcile or unify those divisions which reason also must discursively disassemble" (35). The same defect is said to be operative in the objective spirit as represented in the Hegelian state. According to Habermas, the state is unable, on the face of it, to unify or reassemble the divisions of modern social and political life. Such a "solution" can be assumed only on the supposition of "an absolute which is construed after the model of the self-relation of a cognitive subject" (53). Only when the absolute is conceived as pure or infinite subjectivity, he writes, "can the moments of universality and particularity be thought to be reconciled in the confines of a monological self-knowledge; in the concrete universal (of the state) universal subjectivity thus takes precedence over the individual subject. In the domain of ethical life, this construction yields the priority of the *higher subjectivity of the state* over the subjective freedom of individuals" (53).

The second line of critique takes aim not so much at self-enclosed subjectivity, but at the presumed abstractness or aloofness of Hegelian thought: the tendency of objective and absolute spirit to become the objects of passive contemplation removed from participation in the actual world process. Retired into itself or into its own absoluteness, Hegelian *Vernunft* is claimed to accomplish at best a "*partial* reconciliation"—namely, within the confines of philosophy but divorced from the shared beliefs of public religiosity which, in his early

[6] The applauding comment was quoted above. In a similar vein the study states: "With the concept of the absolute . . . Hegel is able to grasp modernity on the basis of its own principle. In doing so he shows philosophy as the power of synthesis which overcomes all positivities produced by (abstract) reflection" (49). This does not prevent Habermas from claiming that Hegel "misses the essential goal for a self-grounding of modernity: namely, to conceive positivity in such a manner that it can be overcome by relying on the same principle through which it is generated—the principle of subjectivity" (41).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

writings, were meant to link sense and reason, the common people and philosophers. Restricted to its own concerns, Hegelian philosophy—according to Habermas—"ultimately robbed present actuality of its salience, destroyed its intrinsic interest and denied its promise for self-critical renewal" (56). Only latent in his early works, this tendency to passivity is said to surface strongly in Hegel's later system, including his *Philosophy of Right.* At this point, his thought no longer criticized existing reality but only sought to "grasp reality as it is." This "muffling of critique" was a close corollary of the "devaluation of actuality" by philosophy; thus, "the conceptually defined modernity permits the Stoic retreat from itself" (57). As remedy for Hegel's failings, Habermas's study proposes a different model of reconciliation or of the "mediation of the universal and the particular": it is the model of communicative interaction (well-known from his other writings). Instead of subordinating the freedom of individuals to the "higher subjectivity of the state," this model relies on the "*higher intersubjectivity of an uncoerced will formation* within a communication community obeying the need for cooperation" (54). Rather than appeal to the power of *Vernunft,* synthesis here derives from the "universality of an uncoerced consensus achieved between free and equal individuals" (54).

Habermas's reprimands clearly call for a critical response or rejoinder (which I can attempt here only in sharply condensed form). As it seems to me, Habermas's interpretation is lopsided both on the level of certain historical nuances and with respect to key Hegelian concepts. Regarding the former, I find the division between the "young" and the "mature" Hegel—or between a Romantic, mythopoetic outlook and a later pure rationalism—vastly exaggerated, if not "mythopoetic" in turn. In my own reading, Hegel never abandoned his early views on "ethical totality," nor did he dismiss the notions of public religiosity, the "nexus of guilt," or the function of art as emblems of an ethical-social bond; he simply proceeded to reformulate these notions in accordance with the needs of his overall system. Like Plato, he always maintained the correlation of truth, goodness, and beauty, and also the linkage of reason and faith (as is evident in the triad of art, religion, and philosophy on the level of absolute spirit).[7] This leads me to more important conceptual issues.

[7] Regarding public religiosity, one might fruitfully compare Hegel's comments on the topic in *Philosophy of Right* which steer a difficult but fascinating course between established (or "positive") religion and a strict separation of church and state; see *Philosophy of Right*, pp. 165–174 (par. 270).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Habermas chides Hegel's *Vernunft* for remaining locked in a self-contained subjectivity and even in "the confines of monological self-knowledge." Moreover, absolute reason or subjectivity is presented as a "consuming activity of self-discovery" and as an "unconditionally self-productive self-relation." At the same time, however, Hegelian spirit is treated as a detached realm in itself, as a passively contemplated objective reason. Now, clearly, *Vernunft* cannot be both (that is, ceaseless activity and passivity); in my view, it is actually neither. By stressing self-production, Habermas injects into subjectivity a Fichtean flavor of self-constitution. More pointedly: the combination of consuming activity and self-production gives to *Vernunft* a "Young Hegelian" cast—the cast of a "praxis philosophy" relying on self-realization and productivity (which the study elsewhere takes great pains to disavow). On the other hand, the treatment of spirit as an objective realm amenable to contemplation carries overtones of "Right Hegelianism"—a perspective which, as the study indicates, was always bent on transforming *Vernunft* into a set of abstract and heteronomous rational precepts.

To a large extent, Habermas's Hegel chapter thus oscillates precariously between subjective and objective reason, between action and passivity (or else between Left and Right Hegelian vistas). Under the pressure of these opposing trends, Hegel's philosophy is liable to be torn asunder. To restore its unity requires more than patchwork; what is needed, I believe, is an appreciation of the fact that Hegel's "spirit" (like other key concepts) is a metaphysical or ontological category—and not a partisan idea available for direct political utilization. As such a category, "spirit" or "idea" is not simply a subjective capacity (a capacity of self-production) or an objective rational principle, but rather a dimension presupposed by both and in which both are finally reconciled. As is well known, the course of Hegelian philosophy leads from subjective over objective to absolute spirit, or else from simple consciousness over mediated self-consciousness to a consciousness of self-consciousness. On one level, this course is described in the *Phenomenology of Spirit* as a path of "experience"—where experience means neither subjective constitution nor passive endurance. More importantly, the dimension of absolute spirit is not a domain produced or constituted as the outcome of a process of self-production (nor is it externally imposed); rather, it is always already presupposed in the movement of thought, permeating the path of experience from beginning to end. As in the case of all great thinkers, Hegel's philosophizing moved in a circle—not a narrowly self-contained circle but one whose spirals meant to embrace every-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

thing. This aspect of Hegel's thought has been eloquently articulated by Heidegger: "The conclusion of the work has not moved away from its beginning but is a return to it. *The ending is only the transformed beginning which thereby has arrived at itself.* This means, however: the standpoint of the understanding and reenacting reader is from beginning to end, and from the end to the beginning the same—that of absolute reason, of a knowledge which confronts the absolute."[8]

My point here is not to vindicate Hegel's vocabulary or his foundational view of subjectivity. Philosophical developments in our century have amply illustrated the close linkage that always exists between consciousness (or subjectivity) and the unconscious, between enlightenment and processes of darkening or occlusion, between revealment and concealment. My point is simply that no path can possibly be charted beyond Hegel by shortchanging the depth of his insights. Regarding the remedial course proposed in Habermas's study, one can reasonably doubt its viability. In pitting the interest and "subjective freedom" of individuals against the ethical life of the state, Habermas basically invokes Kierkegaardian and Young Hegelian arguments in favor of particularity and concrete individual praxis. From Hegel's perspective, however, individuals removed from public-ethical life are precisely unfree, since freedom is genuinely a public category (and ultimately a synonym for "spirit" or "idea"). By presenting the public sphere as deriving from cooperative "will formation" and the consensus reached between "free and equal" individuals, Habermas's proposal harkens back to the contractarian tradition—a tradition strongly rebuked in *Philosophy of Right*. If the public realm is reduced or subordinated to society (or the sum of associated individuals), Hegel writes, "and if its specific end is defined as the security and protection of property and individual freedom, then the interest of the individuals as such becomes the ultimate end of their association, and hence membership in the state something optional. But the state's relation to the individual is quite different: since the former is objective spirit, it is only as a member that the individual gains concrete objectivity, genuine individuality,

[8] *Hegels Phänomenologie des Geistes, Gesamtausgabe,* XXXII (Frankfurt-Main: Klostermann, 1980), p. 52. Heidegger presents Hegel's philosophy as an "onto-theo-logy," because Hegel's "idea" or "spirit" (*logos*) is also the designation for the essence of being which coincides with God (140/1). The lectures caution against approaching Hegel with "the methods of an *ab ovo* defunct Hegelianism" (121). Compare also Heidegger, *Hegel's Concept of Experience* (New York: Harper & Row, 1970).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and ethical life."[9] To the extent that—to elude contractarian premises—Habermas stresses the rationality of cooperative will formation (reflected in an "ideal speech community"), he merely appeals to a regulative principle which, in Hegel's terms, remains on the level of an abstract "philosophy of reflection." On the other hand, if escape from contractarianism is sought in a concrete "life-world," recourse is taken to the same unreflective or prereflective traditions which were chided in Hegel's early writings. Even when not returning to the *cantus firmus* of Enlightenment rationalism, Habermas's proposal thus retains at best the *disjecta membra* of Hegelian philosophy.[10]

FRED DALLMAYR

University of Notre Dame

ENLIGHTENMENT AND RATIONALITY*

IN modern Europe, enlightening (*aufklärendes*) thought has always been set in opposition to orientations toward the world which are based on tradition. The concept of "Enlightenment"—whether understood as historically designating a movement in thought in the 18th century, or as a concept delineating a process, namely, a trend in the development of the history of civilization—has always been linked to the idea of the overcoming of dogmatic traditions by means of rational insights.[1] Enlightening thought

[9] *Philosophy of Right*, p. 156 (par. 258). While applauding Rousseau's philosophical treatment of the topic, Hegel in the same context criticizes the French thinker for taking "the will only in a determinate form as the individual will," and for regarding "the universal will not as the absolutely rational element in the will, but only as a 'general' will which proceeds out of individual wills as out of conscious wills" (157).

[10] This aspect is revealed in Habermas's general framework (not fully discussed in his Hegel chapter), especially in the juxtaposition of "system" and "life-world." While embracing, on the one hand, a Young Hegelian (and quasicontractarian) view of social praxis and will formation, the framework, on the other hand, makes room for an autonomous system rationality on the level of economy and the state (a rationality largely indebted to Right Hegelian premises).

* To be presented in an APA symposium on Enlightenment and Rationality, December 30, 1987. Fred Dallmayr will be co-symposiast, and Tom Rockmore will comment. See this JOURNAL, this issue, 692–699 and 699–701, respectively, for their contributions.

[1] Herbert Schnädelbach, "Was ist Aufklärung?", manuscript, 1987. (Trans. note: In order to avoid reducing the German word *Aufklärung* to its historical meaning, it has been translated throughout as 'enlightenment', except where reference is specifically to the historical epoch of the same name.)

0022-362X/87/8411/0692$00.70 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

claims to liberate the thinker from the spell of handed-down traditions by subjecting to rational, universally reconstructable examination that which had previously been valid solely by being socially binding. Thus, from the outset, what is unique to enlightenment is its immanent relation to a criterion of rational validity which acts as a standard against which opinions and convictions can be upheld by rational examination. Although it is hardly possible to develop a concept of enlightening thought without such a relation to rationality, ever since the days of that "Enlightenment" epoch, thinkers have nevertheless strongly contested the manner and possibility of providing the foundations of a meaningful concept of human rationality. Even within the innermost circles of the Enlightenment movement, a form of the critique of reason (*Vernunftkritik*) arose—in Rousseau's, Jacobi's, and Herder's writings, and in still sharper form during the early Romantic period—which destined enlightening thought to enlighten itself forever about the concept of reason.[2] Enlightenment has been unfolding ever since as a process of enlightenment about the boundaries and limitations of human rationality; as it feels its way forward, every new step of the critique of reason has been followed by a further stage of the self-restriction of rationality.[3]

Today, this process of a self-critical continuation of enlightenment would appear to have reached an almost insurmountable barrier: with the inadvertent interplay of poststructuralism and neo-Aristotelianism, a broad front has emerged in the critique of reason which applies critique at such a deep level that the project of enlightenment seems to be robbed of its innermost rational medium. I wish to discuss the challenge this poses, first, by distilling the decisive arguments from the complex span of ideas that makes up the new critique of reason and, then, by presenting them in order of decreasing generality and thus of increasing accuracy. I hope thereby to show that, in the process of self-critique which limits the concept of reason accordingly, enlightenment has already found an answer to each respective objection raised by the critique of reason. I shall reconstruct this process in three stages: (1) the critique of reason accused reason of being exclusive in character; this accusation was countered by an anthropological relativization of the position accorded to ra-

[2] Cf. Panajotis Kondylis, *Die Aufklärung im Rahmen des neuzeitlichen Rationalismus* (Stuttgart: Klett-Cotta, 1981); Frederick C. Beiser, *The Fate of Reason: German Philosophy from Kant to Fichte* (Cambridge, Mass.: Harvard, 1987).

[3] Cf. Jürgen Habermas, *Der philosophische Diskurs der Moderne* (Frankfurt am Main: Suhrkamp, 1985).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tionality; (II) the critique of reason objected to the instrumental character of reason; self-critical enlightenment had already parried this objection with a self-differentiation of the categories within the concept of reason; and, finally, (III) the critique of reason has turned on the all-unifying compulsion inherent in a differentiated concept of rationality; this indictment has been answered by the formalistic self-restriction currently adopted by a concept of rationality which stands in the tradition of the Enlightenment. Only after traversing these three completed stages in the self-critique of enlightenment will it become apparent which of the objections raised by the critique of reason have still been left unanswered.

I

At the first, elementary stage of the critique of reason, the view was advanced that any formation of rational attitudes and orientations was ineluctably accompanied by an exclusion of corporeal sensuousness. An exclusive character inhered in rationality or reason, because its use necessitated abstracting from sensuous sentiments and feelings, from needs and predilections. This thesis on the exclusive character of reason found expression in a series of theoretical models. The clearest and, at the same time, most questionable model is to be found in vitalism, which dates back to Schopenhauer and Nietzsche; indeed, some positions adopted in French poststructuralism can be understood as a covert continuation of this train of thought.[4] The objection that reason is exclusive in nature is made sharper by the assertion that rational achievements, in principle, entail an enforced interruption of the normal course of human life and, to this extent, lead to a destruction of corporeal and emotional abilities. Disregarding the fact that the metaphysical implications of such a hypothesis can be criticized without difficulty,[5] this claim lays bare the theoretical premise on which the thesis of the exclusive character of reason must necessarily rely: namely, that the practical process of human life must always be presumed to be an occurrence devoid of rationality; in other words, rational achievements intervene in it as if from outside. For only if rationality is posited as something foreign or diametrically opposed to the practical course of human life can its application be understood as an act of suppression or exclusion of the opposite of reason; if, in contrast, one calls the strict opposition between the sensuous-practical course of life and rational achieve-

[4] Cf. Jacques Bouveresse, *Rationalite et Cynisme* (Paris: Les Editions de Minuit, 1984); Manfred Frank, "The World as Will and Representation," *Telos*, LV (1983): 166ff.

[5] Cf. Herbert Schnädelbach, *Philosophie in Deutschland 1831–1933* (Frankfurt am Main: Suhrkamp, 1983), p. 176ff.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ments into question, then the thesis of exclusiveness loses its suggestive power. Indeed, a theoretical process already set in motion during the age of Enlightenment undermined the opposition of sensuousness and reason by inquiring into the rational content of human feelings and the affective roots of rational achievements.[6] From this starting point, the attempt to desublimate the concept of reason or rationality was then continued by the "Left Hegelians," pragmatism, phenomenology, and philosophical anthropology in such a fashion that gradually rational achievements were being shifted and implanted at an ever deeper level in human life.[7] At the end of this process, rationality was understood not just as a human ability anchored in history, but one also inextricably linked to the body; and, concomitantly, the position of rationality within the whole of human life has been relativized—it is now regarded as that form of reflexive interval which is necessary once a well-functioning form of action hitherto successful in practical terms is disturbed by problems. In this manner, the inclination of subjects capable of action to make use of fallible knowledge reflexively in the solution of problems of action posed in practice comes to be called "rationality." Indeed, this anthropological relativization of reason would seem to have deprived the thesis of the exclusivity of rational achievements of its theoretical basis. For one can no longer assert of a rationality that—as the problem-solving force behind using knowledge reflexively—is completely embedded in the sensuous-practical course of human life that it abstracts from or even destroys a sensuousness linked to the body. The self-critical process of desublimating the concept of reason therefore has, in a certain sense, shifted the coordinates of the Enlightenment's anthropological frame of reference. Reason, as a mental ability, is thus no longer accorded superiority over the process of human life; but, in contrast, as a moment of reflection, it is now so embedded in the course of human life that it serves to solve conflicts or problems that arise therein. Rationality conceptualized in such a manner cannot be accused of destroying or excluding precisely that which, by dint of its own achievements, it is supposed to be recreating, namely, an intact form of the sensuous-practical course of human life.

II

The second, less fundamental stage of the critique of reason centers on the argument that a rational approach is inexorably linked to a perspective that involves the domination of things and persons. Ra-

[6] Cf. Kondylis, *op. cit.*, p. 325ff.
[7] Cf. Habermas, *op. cit.*, ch. XI.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tionality or reason has an instrumental character, because it serves as a means to know (*erkennen*) existing states of affairs and to make it possible to handle practical tasks. This form of the critique of reason—which can take motifs from early Romanticism as precedents and which found impressive expression in *Dialectic of Enlightenment*[8]—starts from the assertion that, by the capacity for rational behavior, we usually mean the ability to know objects in order then to manipulate or control them in terms of the goals of one's own action. It is easy to show that, once it has come to predominate one's dealings with other people or with one's own wishes and needs, such a capacity for purposive-rational knowledge becomes a medium of instrumental domination. Yet, even the age of Enlightenment had itself witnessed the inauguration of attempts both to cast doubt on the one-sided interpretation of reason as a capacity to know objects as well as to prove the existence of other types of "rational" knowledge. Although one already comes across a systematic distinction between different forms of reason in Kant's work, it is, above all, his opponents who endeavor to demonstrate that specific forms of rational attitudes can also guide a person's relation to other human beings and to itself.[9] These initial projects led to the second process of the self-critique of enlightenment, one which began to differentiate internally within the already desublimated concept of rationality by making a distinction between systemically different claims to rationality either made indirectly within the practice of acting or raised implicitly in verbal utterances.[10] For, if "rationality" describes the human capacity for a reflexive mastering of practical problems, then the mere fact that there are different classes of problems which human beings encounter in the course of everyday practice means that a certain compulsion exists to differentiate within the concept of rationality; thus, we claim to behave rationally not only when we encounter a technical problem, but also, for example, if we come into moral conflict with others. As soon as one can demonstrate that the course of human life entails different types of rational knowledge, however, this also highlights the fact that the model of a rational relation to objects which has prevailed in modern Europe constitutes but a one-sided interpretation of the human capacity for

[8] Cf. Max Horkheimer and Theodor W. Adorno, *Die Dialektik der Aufklärung* (Amsterdam: Querido, 1947).

[9] For example, Charles Taylor, *Hegel* (New York: Cambridge, 1975), Part 1.

[10] Cf. Karl-Otto Apel, "Zum Problem der Rationalitätstypen," in *Rationalität: Beiträge zu einer Theorie*, Herbert Schnädelbach, ed. (Frankfurt am Main: Suhrkamp, 1984), p. 15ff.; Jürgen Habermas, *The Theory of Communicative Action*, Vol. I, Thomas McCarthy, trans. (Boston: Beacon 1984), p. 1ff.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

reason. A purpose-directed knowledge of states of affairs, which is what the thesis on instrumentalism suggests is the predominant feature of reason, comprises only one of the different forms that the rational relation of a person to itself and its world can take; therefore, rationality as such can also not be criticized by referring to its instrumental character.[11]

III

The theoretical preconditions of the accusations of logocentrism, which in the shape of the theses on exclusivity and instrumentalism have accompanied enlightenment like a shadow ever since its beginnings, are invalidated by the two stages in the self-differentiation of reason. A rationality which has been relativized with regard to its position in practical life, which has been relinked to corporeal abilities, and which has been embedded in historical praxis and differentiated in different dimensions can with good grounds reject the accusations both of its being, in principle, exclusive and of its necessarily being instrumental. A third stage of the critique of reason, however, one which is particularly influential at present, has charged that, with the emphasis on rational attitudes, the great variety of contents of knowledge and of convictions is inevitably referred back to one single yardstick. Rationality or reason possesses an all-unifying quality, because its claims to universalism force it to make judgments on the validity of statements of different sorts over and beyond all local and temporal boundaries. Whereas the two versions of the critique of modern logocentrism both take the ideal of rationality as such as their point of departure, the third stage in the critique of reason concentrates on the universalistic claims that are ineluctably connected with enlightening thought's reference back to standards of rationality. This claim, raised above all by the representatives of postmodernism,[12] demonstrates that even an enlightenment that has restricted itself via a differentiated concept of rationality nevertheless cannot dispense with universally reconstructable criteria for examining convictions. Yet Kant had already ceased to apply such universalist standards of rationality directly to the contents of individual convictions; he applied them only to the methods whereby they had been produced. Attempts today to effect a formalistic self-

[11] Cf. Herbert Schnädelbach, "Vernunft," in *Philosophie: Ein Grundkurs*, Schnädelbach and Frank Martens, eds. (Hamburg: Rowohlt, 1985), p. 77ff.

[12] Cf. Jean-Francois Lyotard, *The Postmodern Condition: A Report on Knowledge*, George Bennington and Brian Massumi, trans. (Minneapolis, Minn.: Minnesota UP, 1984); for a critique, see my "An Aversion against the Universal: A Commentary on Lyotard's Postmodern Condition," *Theory, Culture & Society*, II, 3 (1985): 147ff.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

restriction of reason move along this path opened up by Kant, implanting the rationality standards of the contents of knowledge in the procedures for their examination: here, rather than in the contents of convictions, only the methods for their examination by argument are still proven to be "rational." This step is also already suggested by the first, elementary process of the self-critique of reason: for, if rationality is understood to signify the capacity of persons capable of action to apply fallible knowledge in solving practical problems, then it makes sense to accept only the rules for such an application of knowledge as the respective standards of rationality. As soon as knowledge is, in addition, conceived of as being mediated through communication, then rationality comes to be measured formally against the procedures of argumentation for the vindication of intersubjectively raised validity claims.[13] Such a step of procedurally self-restricting rationality deprives the charge that reason is inexorably all-unifying of its foundations; indeed, such a charge proves to be a substantive misunderstanding to the extent that "rationality," following the avenue Kant had first opened, is no longer claimed for the contents of individual convictions, but only for the procedures for examining them. Indeed, if the categorial self-differentiation of reason is taken seriously, then different forms of vindication by argumentation must also be shown to exist for the different claims to rationality. Yet, even if it should prove possible to achieve this by taking the promising path of a universal pragmatics pursued nowadays (Karl-Otto Apel, Jürgen Habermas), this in turn means that a new problem arises, namely, that of the formalism innate in such a conception of reason.

It may well be that proceduralism or formalism does indeed constitute the real problem of an enlightenment that has become self-critical. Travelling down the path of the gradual correction of its concept of reason, enlightenment has eventually seen itself forced to restrict its claims to such a great extent that it can of itself no longer generate a substantive knowledge of what is true or false, what is good or bad.[14] A self-critical enlightenment is only in a position to counter indirectly the charge of barren formalism brought against it by neo-Aristotelians, in particular. For it would only be meaningful to rest content with a procedural concept of rationality to the extent that historical or sociological, i.e., empirical proof is forthcoming to show that what enlightenment has identified as rational procedures

[13] Habermas, *The Theory of Communicative Action*, p. 1ff.
[14] For example, Alasdair MacIntyre, *After Virtue* (Notre Dame, Ind.: University Press, 1981), ch. 5.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

for examining forms of problematic knowledge are indeed used within the practices of social life. In such a case, it is admittedly not the philosophical concept of reason itself but rather a life-world that organizes itself rationally which generates a concrete knowledge of that which is true or false, good or bad. Thus, in the final instance, enlightenment will only be able to assert itself if it is able to demonstrate empirically that it itself has, in fact, already become an integral element of social praxis.

AXEL HONNETH

Frankfurt University, West Germany

Translated by Jeremy Gaines

ENLIGHTENMENT AND REASON*

It is no accident that the Enlightenment period, roughly the eighteenth century, is widely known as the Age of Reason. We find here the continued rise of the idea of independent reason which surpasses all forms of superstition, unexamined presuppositions, and dogmatism, especially religious and other arguments from authority. It has been suggested that the movement set in motion by the Protestant Reformation was translated into the philosophical realm through Kant's critical philosophy, which, arguably, culminates in Hegel's system. So Heine, a great poet but also a student of Hegel, argued that in Hegel's thought the circle is closed which began with the rise of the principle of independent reason.

This Hegelian reading of the rise of reason, well known in the writings of the Young Hegelians, has been contested ever since its formulation. Nietzsche, it is well known, held that the very idea of system is indicative of a lack of good judgment, a view which is widely shared by later thinkers. Heidegger, distantly following Nietzsche, saw in Hegel's thought the culmination of the form of metaphysics which represents a fundamental mistake. Jacques Derrida, in Heidegger's wake, has tried to deconstruct the form of thought which depends on the so-called metaphysics of presence.

In his recent writings, the German social theorist Jürgen Habermas has provided a rival reading of the modern concern with reason centered on Hegel and the relativist revolt against Hegel

* Abstract of a paper to be presented in an APA symposium on Enlightenment and Rationality, December 30, 1987, commenting on papers by Fred Dallmayr and Axel Honneth, this JOURNAL, this issue, 682–692 and 692–699, respectively.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

initiated by Nietzsche. His argument can be abbreviated, although not adequately summarized, in the statement that both alternatives are untenable since they remain committed to an exhausted philosophical paradigm, which he calls the philosophy of consciousness (*Bewusstseinsphilosophie*). More precisely, he maintains that the so-called philosophy of consciousness and its relativistic critics are representatives of two dialectically related, but untenable forms of thought. Habermas uses this claim, which he bolsters through extensive historical interpretation, to suggest the need to introduce a supposedly new theoretical paradigm, namely his own theory of communicative action, which he presents as an acceptable alternative to any vestige of the philosophy of consciousness.

In their papers, Axel Honneth and Fred Dallmayr intersect with this analysis on different levels. Honneth regards the Enlightenment as, in principle, opposed to tradition. He sees Habermas's discussion as typical of one phase of the ongoing effort of reason to examine itself in the process of freeing itself from tradition. Dallmayr directs his attention to the accuracy of Habermas's discussion of the main figures in his reading of the modern philosophical tradition, especially Hegel. He maintains that Habermas, in effect, offers a reading of Hegel meant to go beyond the latter's thought which, however, Habermas supposedly does not fully appreciate.

Their papers are complementary, critical contributions to the reception of Habermas's study of reason against the background of the modern philosophical tradition. The virtue of Dallmayr's paper is to test Habermas's reading of post-Kantian philosophy against the positions it is meant to describe. In effect, he calls into question Habermas's reading of Hegel's position for which he proposes another interpretation. In his discussion of Habermas's view as merely another level in reason's self-examination, Honneth suggests that, despite Habermas's radical intent, the latter remains within the ongoing discussion of reason; from this perspective Habermas fails to break with, but only continues the work undertaken by the Enlightenment study of reason.

My criticism of these papers is not that they go too far; rather, I believe that they do not go far enough either in the description or the critique of Habermas's overall project. That project, I have argued elsewhere, can be regarded as a quasi-Hegelian effort to interpret the history of modern philosophy and to overcome its alleged shortcomings in a view of reason as socially relevant and critical in a Kantian sense. Habermas's intent, then, is to fulfill the intrinsic aim of modern philosophy, especially German idealism and above all

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Marx's thought, from the perspective of another theoretical paradigm, namely his own theory of communicative action.

If this is correct, then the defects in the two discussions of Habermas's position are symmetrical. Dallmayr, who concentrates on the reading of the history of philosophy, does not address Habermas's notion of reason which underlies it. Honneth, who concentrates on Habermas's idea of reason, does not deal with the analysis of the history of philosophy meant to justify it. Neither critic directly considers what I regard as the central problem in Habermas's analysis of reason, namely, whether reason in the most general sense is socially relevant. Or, to put this same question in terms of Habermas's own extended debate over many years with historical materialism, we need to inquire whether it is possible to fulfill the intention of Marx's position through another, intentionally incompatible theoretical paradigm.

TOM ROCKMORE

Duquesne University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

The editors report with sorrow the death of Irving G. Thalberg, Jr., Professor of Philosophy at the University of Illinois at Chicago, on August 21, 1987 in Syracuse, NY. Professor Thalberg received his Ph.D. from Stanford University in 1960; he taught at Oberlin College from 1960 to 1963 and at the University of Illinois since 1965. At the time of his death he was 56 years old.

The Australasian Association of Philosophy and the Australasian Association of Logic will hold joint conferences on May 15–20, 1988 at the University of Western Australia. The conferences will include sessions on the history of Australian philosophy, women in philosophy, cognitive science, and Hobbes; an official call for papers will be issued shortly, with a probable deadline of February 1988. For more information, please write to: Dr. Graham Priest, Department of Philosophy, University of Western Australia, Nedlands 6009, Australia.

The Bertrand Russell Society announces a call for papers to be presented at its meeting with the Eastern Division of the American Philosophical Association in December 1988. Papers may be on any aspect of Russell's philosophy. The submission deadline is April 1, 1988. For more information, please write to: David E. Johnson, Chairman, Philosophers' Committee, The Bertrand Russell Society, Sampson Hall, U.S. Naval Academy, Annapolis, MD 21402-5044.

The London School of Economics and Political Science is seeking nominations for the 1988 Lakatos Award in Philosophy of Science. The Award, which has been endowed by the Latsis Foundation in memory of Imre Lakatos, will be for an outstanding book on the philosophy of science published in English during the last ten years (i.e., in 1978 or later). The closing date for nominations is April 15, 1988. For more information, please write to: The Secretary, The London School of Economics and Political Science, Houghton Street, London WC2A 2AE, England.

The recently established Jerusalem International Spinoza Institute announces a series of seven conferences entitled "Spinoza by 2000: The Jerusalem Conferences." The first conference in the series, entitled "Ethica I: God and Nature—Spinoza's Metaphysics," was held last April. The next conference, "Ethica II: the Human Mind and the Theory of Knowledge," is scheduled for 1989; conferences will be held every two years thereafter through 1999. For more information, please write to: The Jerusalem Spinoza Institute, Mishkenot Sha'ananim, Post Office Box 8215, Jersualem 91081, Israel.

The Israel Academy of Sciences and Humanities and Mishkenot Sha'ananim announce an "International Conference on Justice in Punishment: Philosophical, Legal, and Criminological Issues Involved in the Practice of Punishment," to be held in Jerusalem on March 27–31, 1988. The main topics

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

will be: theories of justice and the problem of punishment; theories of punishment; capital punishment; determining penalties; and alternatives to punishment. For more information, please write to: Justice Conference Secretariat, Mishkenot Sha'ananim, Post Office Box 8215, Jerusalem 91081, Israel.

Philosophers who are planning a visit to Israel and who would like to deliver lectures or meet other philosophers are encouraged to get in touch with the Israel Philosophical Association, which will be glad to be of help in these matters. For more information, please write to: Asa Kasher, President, Israel Philosophical Association, Department of Philosophy, Tel-Aviv University, Tel-Aviv 69978, Israel.

The North American Chapter of the Society for Greek Political Thought invites all political scientists, philosophers, classicists, historians, and specialists in comparative literature interested in scholarly interchange to join its Society, which includes over 300 members world-wide involved in the study of classical political thought, broadly defined. For more information, please write to: Dr. Kent F. Moors, Director, North American Chapter, Society for Greek Political Thought, Department of Political Science, Duquesne University, Pittsburgh, PA 15282.

The History and Philosophy of Science Section of the American Association for the Advancement of Science will sponsor two symposia at the next annual meeting of the Association at the Prudential Center in Boston, February 11–15, 1988. The first symposium, entitled "The Role of Hypotheses in Science," on Saturday, February 13, will include Peter Achinstein, John Earman, Abner Shimony, and Lawrence Sklar. The second, on Monday, February 15, will be entitled "Science, Engineering, and Ethics: The State of the Art"; participants will include Rachelle Hollander, Deborah Johnson, Roger Kasperson, Sheldon Krimsky, Patricia Rose, Pam Seidenman, Joan Sieber, Susan Snyder, Nicholas Steneck, Albert Teich, and Patricia Woolf. For more information, please write to: Peter Achinstein, Department of Philosophy, Johns Hopkins University, Baltimore, MD 21218.

The New York Academy of Sciences and the Philosophy and Technology Studies Center of Polytechnic University, Brooklyn, NY, announce a conference entitled "Ethical Issues in Military Research: What Are the Issues? How Can They Be Handled?" to be held at the New York Academy of Sciences, May 18–20, 1988. The purpose of the conference will be to bring scientists and engineers from both sides of the political-moral debate over the Strategic Defense Initiative (SDI) together with philosophers and others who can help place such debates within a larger context. Issues to be considered will be: conceptual distinctions between pure and mission-oriented research; the historical and sociological background of current debates; the kinds of ethical arguments employed in the articulations of different positions; and the moral options open to persons on various sides of the issues. For more information,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

please write to: Conference Department, New York Academy of Sciences, 2 E. 63rd Street, New York, NY 10021.

Columbia University is pleased to announce the James Gutmann prize in the Humanities, to be given annually to a Columbia College graduating senior who has made a distinguished record in the humanities or who has shown special insight into humanistic values. The prize was endowed by the family and friends of James Gutmann, Professor of Philosophy Emeritus, on the occasion of his 90th birthday. Dr. Gutmann has been a member of the Columbia philosophy department since 1920, and was chairman from 1953–60. He wrote articles and book reviews for this JOURNAL from 1919 to the late 1970s.

The Department of Philosophy of Washington University/St. Louis is pleased to sponsor an International Conference on Quine, to be held April 9–13, 1988. Participants include A. J. Ayer, Lars Bergstrom (Uppsala), J. Richard Creath (Saint Xavier), Donald Davidson (UC/Berkeley), Burton Dreben (Harvard), Hartry Field (USC/LA), Dagfinn Føllesdal (Oslo), Paul Gochet, Susan Haack (Warwick), Ian Hacking (Toronto), Gilbert Harman (Princeton), Jaakko Hintikka (Helsinki), Jerrold Katz (CUNY), Dirk Koppelberg, Henri Lauener (Bern), Michele Leonelli (Univ. Degli Studi di Pisa), Ruth Barcan Marcus (Yale), Alex Orenstein (CUNY), Charles Parsons (Columbia), W. V. Quine (Harvard), Anthony Quinton, P. F. Strawson (Oxford), Barry Stroud (UC/Berkeley), Joseph Ullian (Washington/St. Louis), and Hao Wang (Fu Jen Catholic). For more information, please write to: Quine Conference, Department of Philosophy, Campus Box 1073, Washington University, Saint Louis, MO 63130.

The 1988 conference of the Academy for Jewish Philosophy, sponsored by the Religion Department of Temple University and the Reconstructionist Rabbinical College, will take place on May 15–16, 1988, in Philadelphia. The theme of the conference will be "Judaism and Hermeneutics," and participants will include: David Bleich (Yeshiva), Eugene Borowitz (H.U.C.-J.I.R.), Joseph P. Cohen (St. John's), Elliot Dorff (Judaism), Seymour Feldman (Rutgers), Marvin Fox (Brandeis), Daniel H. Frank (Kentucky), Martin Golding (Duke), Bernard Goldstein (Pittsburgh), Arthur Hyman (Yeshiva), Alfred L. Ivry (Brandeis), Steven Katz (Cornell), William Kluback (Kingsboro), Barry Kogan (H.U.C.-J.I.R.), Jonathan Malino (Guilford), David Novak (Baruch), Peter Ochs (Colgate), Norbert Samuelson (Temple), Steven Schwarzschild (Washington), Susan E. Shapiro (Temple), David Shatz (Yeshiva), Kenneth Seekin (Northwestern), David Silverman (Philadelphia), Jacob Straub (R.R.C.), and Josef Stern (Chicago). For more information, please write to: Norbert M. Samuelson, Religion Department, Temple University, Philadelphia, PA 19122.

The Philosophy Department at Virginia Polytechnic Institute announces a conference entitled "Methodological Approaches to Plato and His Dia-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

logues" to be held April 21–24, 1988. There will be twelve lead speakers and ten commentators: Julia Annas (Arizona), Mary Whitlock Blundell (Classics, Washington), Thomas Brickhouse (Lynchburg), Diskin Clay (Johns Hopkins), William Cobb (William & Mary), Dan Devereaux (Virginia), Gail Fine (Cornell), Michael Frede (Princeton), David Halperin (Classics, M.I.T.), Charles Kahn (Pennsylvania), David Keyt (Washington), George Klosko (Government, Virginia), L. A. Kosman (Haverford), Edward Lee (California, San Diego), Mark McPherran (Maine at Farmington), Martha Nussbaum (Classics & Comp. Lit., Brown), Cynthia Patterson (History, Emory), Richard Patterson (Emory), Arlene Saxonhouse (Politics, Michigan), Kenneth Sayre (Notre Dame), Kenneth Seeskin (Northwestern), and Henry Telo (Vanderbilt). For more information, please write to: James Klaffe and Nicholas Smith, Department of Philosophy, Virginia Polytechnic Institute, Blacksburg, VA 24061.

The Department of Philosophy of Elmhurst College, Elmhurst, Illinois is pleased to announce the appointment of James S. Kelly as Assistant Professor, effective September 1987.

The Department of Philosophy of the University of Minnesota/Duluth, announces the appointments of James Fetzer as Professor of Philosophy and Michael Losonsky as Visiting Assistant Professor of Philosophy, effective September 1, 1987.

The Philosophy Department of the University of New Mexico is pleased to announce that Abraham Anderson (Columbia), Rose-Mary Sargent (Notre Dame), and John Taber (Hamburg) have been appointed as Assistant Professors of Philosophy, effective September 1987.

The Department of Philosophy of the University of Massachusetts/Amherst is pleased to announce the appointment of Donald Morrison as Visiting Assistant Professor, effective September 1, 1987.

The Department of Philosophy of Denison University is pleased to announce the appointment of Dr. Babette Babich as Assistant Professor of Philosophy, effective September 1987.

The Center for Philosophy of Science at the University of Pittsburgh is pleased to announce that William Bechtel (Georgia State), Robert Butts (Western Ontario), Richard Healey (U.C.L.A.), Anne Hiskes (Connecticut), Paul Hoyningen (Zurich), Peter Janich (Marburg), Ulrich Majer (Hannover), Nicholas Maxwell (University College, London), Victor Rodriguez (Observatoria Astronomico, Argentina), Stewart Shapiro (Ohio State, Newark), Barbara Tuchanska (Lodz, Poland), and Danilo Zolo (Florence) will be visiting fellows during the 1987–88 academic year.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

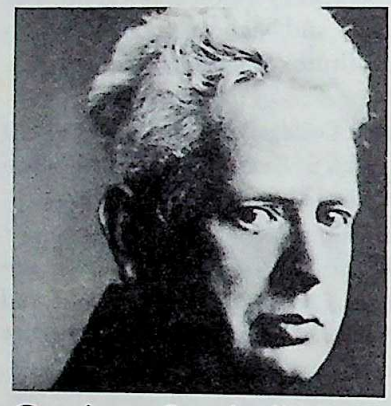
Digitized by Arya Samaj Foundation Chennai and eGangotri

Yale

Woman, Nature, and Psyche
Patricia Jagentowicz Mills

A critical analysis of the relationship between the domination of nature and the domination of women in society.
$26.50

Desire, Dialectic, and Otherness
An Essay on Origins
William Desmond $29.50



Cassirer: Symbolic Forms and History
John Michael Krois $25.00

Heidegger and the Philosophy of Mind
Frederick A. Olafson $23.50

Marcus Aurelius
A Biography
Revised Edition
Anthony Birley $25.00

Habermas and the Dialectic of Reason
David Ingram $24.50

Electric Language
A Philosophical Study of Word Processing
Michael Heim

Heim provides the first consistent philosophical basis for critically evaluating the impact of word processing on our use of and ideas about language.
$19.95

Novalis: A Romantic's Theory of Language and Poetry
Kristin Pfefferkorn

This innovative view of Novalis's life and work focuses on his important contributions to the philosophy of language—particularly the relation of language to time and space. $30.00

Slightly Beyond Skepticism
Social Science and the Search for Morality
Leonard Doob

"Doob offers a unique angle on philosophical questions, an angle that brings some issues into clearer focus and that will enrich the work of moral and political philosophers." —James S. Fishkin
$25.00

Doctrines of Religious Communities
A Philosophical Study
William A. Christian, Sr. $25.00

At the Origins of Modern Atheism
Michael J. Buckley, S.J. $27.50

Explaining Religion
Criticism and Theory from Bodin to Freud
J. Samuel Preus $25.00

Nietzsche's Teaching
An Interpretation of Thus Spoke Zarathustra
Laurence Lampert $29.50

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXIV, NO. 12, DECEMBER 1987

19131888

गुरुकुल कांगड़ी विश्वविद्यालय हरिद्वार

SERIOUS PHILOSOPHY AND FREEDOM OF SPIRIT*

PHILOSOPHERS have long taken sides in a great divide by adopting one or more of a set of views that tend to go together:

FREEDOM	SERIOUSNESS
Subjectivism	Objectivism
Relativism	Absolutism
Antirealism	Realism
Historicism	Universalism

On the left is "free-spirited philosophy"; on the right "serious philosophy."

I wish to lay out a view of seriousness, and to consider recent attacks on that view from the side of the free spirited: deconstruction and textualism, hermeneutics, critical theory, and the new pragmatism. Without defining what all forms of freedom have in common, I shall draw from them a combined critique against seriousness. I will also examine, occasionally and in passing, positive ideas conjured up by the free. But mainly I wish to consider their combined critique of seriousness.

The combined critique—labeled here "free spirit" for convenience—is not one that *all* "free spirits" would accept. But all sides do contribute, some more extensively than others. Moreover, the

* Earlier versions were presented as public lectures at Williams College and Muhlenberg College; at international conferences in Spain and Mexico; and as a plenary address at the World Congress of September 1987, in Argentina. It benefitted from the discussion elicited, and from the written comments of many colleagues.

0022-362X/87/8412/0707$02.00 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

influence of free spirits extends far beyond academic philosophy. It is found not only in historical studies, with Michel Foucault, but also in literary and critical theory, now widespread, with Jacques Derrida, Richard Rorty, and less serious cohorts. An allied phenomenon in legal scholarship is the critical legal studies movement, whose theorist Roberto Unger[1] preaches the sway of ideology over all value and norm, and the optionality and groundlessness of all politico-legal frameworks; and also advocates constant change and revision—framework shattering (groundlessly for its own sake, presumably).

I. SERIOUS PHILOSOPHY

Serious philosophy here means the Platonic tradition committed to an objective reality knowable to mind but independent from being known or conceived or otherwise minded—independent anyhow from its *human* minding. This makes room for idealism, and allows even Berkeley his rightful place in that tradition. Berkeleyan spirits, from God on down, really do their minding without that minding requiring for its reality the support of any minding directed upon it; in that sense, therefore, it has a reality independent of being minded. The same is true for the Cartesian cogito, and for much axiology or ethics whose foundation for the good or the right resides in human preference and allows that goodness and rightness such objectivity as may lie in the reality of such preference—a preference which might exist in itself and unminded.

According to serious philosophy, moreover, objective reality provides us not only with objects of knowledge, but also with constraining starting points and obligatory methods for reason. These ultimate starting points and ultimate methods define an ahistorical framework that determines the legitimate use of reason and fixes the conditions and extent of rationality, knowledge, goodness, and rightness.

Some would think such philosophy positively *dour*. One or another element would be denied not only by free spirits but by many who view themselves as serious but not dour.[2] For example, many would reject moral realism and evaluative objectivity generally, while

[1] "The Critical Legal Studies Movement," *Harvard Law Review*, XCVI (1983); and *Passion: An Essay on Personality* (New York: The Free Press, 1984).

[2] Even Habermasian critical theory is a special case. It is very hard to place, partly because it vacillates on whether to go transcendental. As Raymond Geuss shows, Jürgen Habermas changes his mind already in the mid 1960s; see *The Idea of a Critical Theory* (New York: Cambridge, 1981), p. 64: "In some of his earliest essays Habermas follows Adorno and holds a contextualist view of reflection; then, sometime in the mid 1960s, he seems to have been frightened by the specter of relativism, and retreated into a kind of transcendentalism." This is then supported textually.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

accepting objective facts of nature or mathematics. Others would reject realism wholesale with regard to values, nature, *and* mathematics. Logical positivists once rejected all realism by rejecting the very question of realism as a worthless pseudoquestion. But the question has proved hardier and more intriguing than its detractors. Today realism is again widely debated across the whole field of philosophy, and even moral realism once more elicits serious attention.

Serious philosophy and its free-spirited opposition lie at opposite ends of a spectrum, therefore, and no *pure* exemplar of either may be accepted by anyone today. I have chosen to compare extremes for the sake of a starker, more dramatic contrast. And, for convenience, I shall classify a philosophy as free spirited if it contains some significant admixture of the pure article.

"Is there a single major school or movement of philosophy—analytic or continental—which today approaches the serious end of our spectrum?" Suppose there is not. Why should one care? Schools and movements attract the public eye by dramatic opposition to the conventional wisdom. The latter, by contrast, is seldom unified into any explicit, concise formulation, and even more seldom is it baptized with any catchy label.

The present paper may be viewed as pitting analytic philosophy against all other philosophy. But that is a twofold mistake. Too much of analytic philosophy is not serious; think of the varieties of neopragmatism, only the most abandoned of which is considered explicitly here. And too much of all other philosophy is not free; think of institutional Thomism or Marxism. Free-spirit philosophy is certainly represented within analysis, as is serious philosophy outside analysis. Our lines of controversy are *not* drawn geographically or ideologically.

A final preliminary caveat: no proof is offered here for the tradition of seriousness, nor even weighty reasons in its favor. I wish rather to consider reasoning aimed in recent years *against* that tradition.

This introductory section has made the following six points:

(a) Serious philosophy encompasses Platonic traditions of objective realism.
(b) It believes in real objects of knowledge and in objective methods for reason.
(c) Seriousness and freedom lie at opposite ends of a spectrum; most of us fall somewhere in the middle.
(d) Perhaps no celebrated school or movement lies today near the serious end—but that is of limited importance.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(e) The division between seriousness and freedom is not geographical or ideological.

(f) No positive argument for seriousness shall be offered here—only a defense against attacks from the side of freedom.

II. PHILOSOPHY FOR FREE SPIRITS

It is a main traditional goal of serious philosophy to uncover more and more of the objective framework of reason. What can possibly remain for philosophy to do if we reject the very existence of such a framework? What remains is, of course, the task of free-spirit philosophy, variegated as it is bound to be. In the book of some free spirits, the Word was not only in the beginning, but is even—incredibly—said to be *everything*.[3] Free spirits are often textualists, readers and authors who live for the conversation, the only point of which, insofar as it has any extrinsic point at all, is to go on and on without end. Free-spirit philosophers do try to understand how things in the broadest sense of the term hang together in the broadest sense of the expression. And they claim to be aided in that endeavor by their lighter metaphysical baggage and their less ambitious epistemological commitments. Such free spirits would function in society as all-purpose critics or utility intellectuals, renouncing any shared disciplinary matrix and any common professional purpose.

So much for a life and direction found more lovely and apt by free spirits of our day. Here now is their brief against seriousness.

III. OBJECTIONS TO THE TRADITION OF SERIOUSNESS

It has been charged that only an unwarranted fear supports the tradition of serious philosophy, a threatening either/or: either there is the fixed, binding framework of reason, or we face intellectual and moral arbitrariness and chaos. According to some, it is this Cartesian anxiety[4] that needs to be exorcised. Once emancipated from that dichotomy, we may then move beyond objectivism and relativism, a task more practical than theoretical. What the world needs now is a profound recognition that the proper end of man is not to mirror

[3] According to Derrida, "there is nothing outside the text," and "there has never been anything but writing." *Of Grammatology*, G. C. Spivak, trans. (Baltimore: Johns Hopkins University Press, 1976), pp. 158/9. Derrida's pronouncements find an echo in Rorty's introduction to his *Consequences of Pragmatism* (Minneapolis: University of Minnesota Press, 1982), p. xxxvii: "The intuitive realist thinks that there is such a thing as Philosophical truth because he thinks that, deep down beneath all the texts, there is something which is not just one more text but that to which various texts are trying to be 'adequate'. The pragmatist does not think that there is anything like that."

[4] In Richard Bernstein's apt phrase: *Beyond Objectivism and Relativism* (Philadelphia: University of Pennsylvania Press, 1983), p. 18.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

passively any dead objective reality, but to work actively for solidarity among the living.[5]

Two chief arguments urged by free spirits deserve serious attention. The first is an argument against theoria and for phronesis, against the possibility of algorithmic calculation on any question of values, be it practical or theoretical, and in favor of deliberation and wise judgment. The second flows from the alleged failure of serious philosophy. "Given its total lack of success," we are told, "it is time to put a stop to its pretensions. Epistemology in particular deserves to be laid to rest, for its oft-repeated lie that it would found culture if allowed the time and attention." We consider these two arguments in turn.

IV. FIRST ARGUMENT AGAINST SERIOUSNESS: PHRONESIS VERSUS THEORIA

Invective now converges from all sides on the project of legitimizing or founding segments of our culture. For example, we are told that justice and freedom are incommensurable, and that there is no true principle that can give general priority to either over the other, nor is there any that will show us in advance how to resolve all possible conflicts, nor even all likely or actual ones. It is also well known that the judge and the legislator receive from ethics or morality no universal principles from which it might always be possible to deduce the right verdict or legislation. Deliberation requires, on the contrary, the pondering of reasons in favor and against, and the estimation of their respective combined weights as well as an assessment of the balance. All of that seems to me both true and important. If to found or legitimize requires providing axioms or even just coherent and general theories that would logically imply all that is to be founded, humanity is obviously in no position to "found" either law or morality or science. And, if we did manage to provide such a foundation for any of our science, that would just be to succeed with more and better science. One thing should be clear, nevertheless, for science as well as for law and morality: even if our success to date has been modest, that does not show we ought to quit. Where would physics be if Galileo, Newton, and Einstein had yielded to such reasoning? *Besides*, suppose that *fundamentally* there are *several* incommensurable factors; and suppose it to be quite *impossible* to arrive at a

[5] Compare Rorty, "Solidarity or Objectivity?" *Nanzan Review of American Studies*, VI (1984). The theme of a philosophy of solidarity is drawn by Bernstein from the writings of Hannah Arendt, Gadamer, Habermas, and Rorty. See "Praxis, Practical Discourse, and Judgment," *Consequences*, pp. 171–233.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

general and coherent theory that will logically imply a decision for all issues which might possibly or even probably come before us. Even so, it does not follow we ought to find for freedom and against seriousness. Fundamental factors that weigh for or against when one evaluates or deliberates may after all be factors whose validity, though incommensurable, is yet objective, absolute, and universal. Each fundamental factor, whether positive or negative, may have its own vectorial character objectively, absolutely, and universally.

So much in regard to the values and reasons that enter deliberation and evaluation. Let us next consider more generally the nature and validity of our beliefs.

"Accuracy, consistency, scope, simplicity, and fruitfulness" are values that, according to Thomas Kuhn,[6] inform scientific theory choice. When these values are in conflict, says Kuhn, rational scientists may still disagree in their judgments without showing any lack of information or failure of intellect. What Kuhn has claimed about science is found by Hans-Georg Gadamer (with inspiration from Aristotle) all over our lives and culture, and is accordingly made a major theme of his philosophy over nearly the full stretch of a long career. Deliberation and practical judgment are different in kind from mechanical proof or algorithmic calculation. A proper model is the judge invoking precedents, and not the accountant doing his sums. If you solemnly promised to be on time for a meeting but then your child needs some medicine and might be helped by getting it early, and you cannot both be on time and give the medicine early, what are you to do? Is there any rule to recall from mother's knee, or draw from religion, or get from our shared fund of wisdom, to solve our problem algorithmically? Can we even hope to *invent* a rule with any plausibility at all? Not when crucial factors seem in the relevant sense incommensurable. How solemn was the promise? How important is the meeting? How weighty is the claim of the group on one's loyalty? How serious is the child's illness? What are the possible consequences? What risks or discomforts are implied by delaying? How assured can we be of these matters? However assured we may be, there is simply no hope for algorithmic calculation on our choice as parent. The factors involved do not severally admit each its own metric, much less a common metric that would permit calculation—none accessible to us, anyhow, either now or in our most remote horizon. In such matters, we simply have no substitute for delibera-

[6] "Objectivity, Value Judgment, and Theory Choice," in *The Essential Tension* (Chicago: University Press, 1977).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
tion and choice by a wise and sensitive judge, no substitute for phronesis rather than theoria.

Even *if* we grant that our lack is not merely epistemic, that no increase in intelligence or information would help (which seems less than certain and obvious), what follows? Just because reason is not in a straightjacket, do we infer it can jump to the moon? That would be lunacy indeed. We have long known that reality is not determinate, not at least in terms accessible to our limited brains. When does night begin, or even twilight? When does baldness begin, or even thinness of hair? Such questions are to be met at every turn, whether we notice them or not. In the absence of any sharp difference between baldness and nonbaldness, there is yet the clear difference between a shiny pate and a full head of hair. Day turns into night at no precise instant, but noon differs *toto caelo* from midnight.

Accordingly, if the Cartesian anxiety is the fearsome either/or that requires algorithmic determinacy on pain of chaos, it might be possible to exorcise it by reflecting on how obvious it is that right now nearly everywhere we find neither the determinacy nor the chaos. From the fact that no one alternative is forced objectively, it is a fallacy to infer that no alternative is objectively foreclosed. And, if Buridan's ass is regarded as absurd and his predicament repugnant to reason, that may call for edifying therapy, but which among us is today so radical a devotee of reason?[7]

Naturally, the application of phronesis in particular situations cannot take place in a logical vacuum. Deliberative reason needs pertinent values to weigh, and factual assumptions in whose light to apply them. Where can the free spirit turn for such values and assumptions? Why does one accept that a promise was made or that a child is ill? Why suppose that such factors are relevant to one's decision? Gadamer speaks for many when he appeals here to the lore of the tribe, to tradition and precedent. And it is, of course, undeniable that we are creatures of our time and group, schooled in its way of life and informed by its habits of thought from the dawn of our reason. But our question requires not sociology or anthropology, but philosophy. We wish to know *not only* how we get to be the way we are. We wish to know whether and why it is good or right or fitting

[7] Martha Nussbaum also draws inspiration from Aristotle for her theme of the place and nature of phronesis. Important here are her recent paper, "The Discernment of Perception: An Aristotelian Conception of Private and Public Rationality," *Proceedings of the Boston Area Colloquium in Ancient Philosophy*, 1 (1985): 151–201; and her book, *Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy* (New York: Cambridge, 1985).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

for us to be that way. We wish to know not only how we get to believe that a high thermometer reading means a high fever and a high fever an illness. We wish to know also whether and why it is that deriving our belief that way makes it an appropriate thing to believe. We wish to know not only how we get to believe that having promised is a good reason for keeping one's promise, but also whether and why that is a good thing to believe.

Getting to know how we acquire our commonsensical or scientific beliefs does perhaps give us a large part of the answer to philosophical questions about the nature and justification of our beliefs. But it cannot possibly be the whole answer, since plenty of superstitious and evil beliefs also get acquired all the time, and they too admit sociological or anthropological explanation. Plenty of evil and superstitious beliefs get acquired by imbibing tradition and the lore of a tribe. Nazi ideology, South African apartheid, and the mountain people's form of life are only three notorious examples.[8] So it is natural for us to wonder what the difference is, to wonder what we can say most generally to explain what makes our way better than theirs (in *some* respects anyhow), and our body of beliefs more credible. It might be difficult to generalize, but, if there really are general differences, as undeniably there are, it should be possible to state them and give them some degree of systematic order, and to provide answers with some degree of generality about the superiority of one system over the others. It is true that, at some level, there will be no real alternative to simply reaffirming the content of one's own system and hence begging the question *if* one is engaged in debate. But that does not mean one necessarily acts inappropriately in doing so, any more than reaffirming the law of contradiction would be inappropriate in the face of its denial. Some questions demand begging. This response has been called "too easy," but it is hard to fathom what that could mean. Suppose we reach a position to say that, at bottom, it is always factor *F* that makes a belief of our own justified, whatever the belief might be. *F* might be thought to be experience, or intuition, or convention, or pragmatic working, or coping, or coherence, or perhaps a combination of logic and observation—if we are lucky it will be something simple and unified. And suppose we are then challenged to explain what justifies us in holding that very belief. What can we possibly say in reply except that here again it is the presence of factor *F*? Will someone object that we beg the question and argue in a circle? The answer is, of course, that we do, but that,

[8] All three are tarred, though not with the same brush. Cf. Colin Turnbull's *The Mountain People* (New York: Simon and Schuster, 1972).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

at this level of generality, it simply cannot be helped, and that regretting it is as bad as regretting that we cannot make a round cube.[9]

At just this point in the argument, there is a great temptation to say that, fundamentally, it is all relative and that factor *F* is just *our* basic source of justification, that others may differ, and, if so, to each his own. And this trend, once begun, extends easily to truth and morality as well. Now, if the claim is that there is no truth, justification, goodness, or rightness, *period,* but only justification, etc., *relative* to system *Si* (for *S*1, *S*2, . . .), then our justification for our own system is just relative to itself, and that is just to say that it affirms its own justification, truth, or what have you, that it affirms this at least implicitly, by implication. But, of course, many radically different and seemingly clashing systems can do the same, including that of the Nazis, that of the present South-African leadership, and perhaps even that of the mountain people. Whence the need to burden our own system with an axiom that there is no justification (truth, etc.) except relative to a system, in the sense that there is not even the very notion of justification (truth, etc.) on its own, *period?* That is just self-destructive and unnecessarily so. If all I say when I say that our system is justified (etc.) is that it is justified (etc.) relative to itself, and, if all this amounts to is that it approves of itself, either explicitly or implicitly, and, if many clashing systems do the same, how then can my commitment to our system be anything more than arbitrary willfulness? And, if that is its status, how can the justification of anything relative to such a system rise above that level of arbitrary willfulness? Where would the extra increment of justification come from? And how could it possibly make up the lack in its fundamental principles, the principles relative to which one must attain whatever justification one ever does attain?

Gadamer would offer the following defense:

> It is a grave misunderstanding to assume that emphasis on the essential factor of tradition which enters into all understanding implies an uncritical acceptance of tradition and sociopolitical conservatism. . . . In

[9] That reasons inevitably give out is a familiar theme of the later Wittgenstein. But does that impose relativity to a form of life? And, if there is no way to understand how a private party could be wrong except by reference to a public form of life, what is the gain really if there is still no way to understand how the *group* could be wrong? Or shall we postulate a group of groups? And where do we appeal above the highest metagroup? Could this be a new argument for God, or is it rather the same old idea of God as the necessary ever-present thinker of ideas? See Kripke's *Wittgenstein on Rules and Private Language* (Cambridge, Mass.: Harvard, 1982), for reflections (and variations?) on the theme from Wittgenstein.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

truth the confrontation of our historic tradition is always a critical challenge of this tradition. . . . Every experience is such a confrontation.[10]

But, as Richard Bernstein argues in response, we are left wondering what one might appeal to in a critique of one's own tradition:

> The claim that Gadamer makes about Aristotle's understanding of natural law (the universal element) as something that is essentially open to interpretation and that is only specified when mediated in a concrete ethical situation that demands choice and decision is paradigmatic, for Gadamer, of the application of all ethical principles and norms. But as Aristotle stresses, and Gadamer realizes, *phronesis* presupposes the existence of *nomoi* (funded laws) in the *polis* or community. This is what keeps *phronesis* from degenerating into the mere cleverness or calculation that characterizes the *deinos* (the clever person). Given a community in which there is a living, shared acceptance of ethical principles and norms, then *phronesis* as the mediation of such universals in particular situations makes good sense.
>
> The problem for us today, the chief characteristic of our hermeneutical situation, is that we are in a state of great confusion and uncertainty (some might even say chaos) about what norms or "universals" ought to govern our practical lives. Gadamer realizes—but I do not think he squarely faces the issues that it raises—that we are living in a time when the very conditions required for the exercise of *phronesis*—the shared acceptance and stability of universal principles and laws—are themselves threatened (or do not exist). . . .
>
> Furthermore, Gadamer does not adequately clarify the type of discourse that is appropriate when questions about the validity of basic norms (universals) are raised. How is such recognition and agreement to be achieved? When there is serious disagreement about what norms ought to be binding, should all participants be able to have a say? When pressed about these questions, Gadamer frequently deals with a different issue. He tells us that such universals are inherited from tradition, that they are essentially open, that their meaning can be specified only in application to concrete, practical situations. But this does not clarify the issue of what we are to do in a situation in which there is confusion or conflict about which norms or universals are appropriate, or how we are to evaluate a situation in which we question the validity of such norms (*op. cit.*, pp. 154–158).

If phronesis is, by definition, the application of laws already accepted in the community, it is of no use in deciding an issue for which there are no pertinent laws, which in a given community might even in-

[10] "The Problem of Historical Consciousness," in Paul Rabinow and William M. Sullivan, eds., *Interpretive Social Science: A Reader* (Berkeley, Calif.: University Press, 1979), p. 108.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

clude the issue of how to overcome some perceived defect in a basic law. Moreover, if phronesis is so defined, how can one even make sense of the very notion of a defect in a basic law?

Is there not indeed a deeper phronesis which is not to be defined as the application of already accepted rules or laws, but is rather to be employed in the assessment of even the most fundamental community *mores?* Is there not deliberation, weighing of relevant values and reasons, even on matters untouched by such mores, or at least on matters that admit plenty of factors whose pertinence does not derive from any commonly accepted rule or law?

What is more, and what seems crucial here, are there not two fundamentally different sorts of laws? There are laws whose ultimate legitimacy and foundation lie simply in community agreement, so that, if the community had opted for requiring the opposite behavior, then that would have been the required behavior in the circumstances, with just as much legitimacy and foundation. For example, whether we should drive on the right or, on the contrary, should drive on the left depends simply on what rule the community adopts. But whether one should kill someone else at random and in cold blood just for the fun of it does not seem the same sort of question. Refraining from such activity is not a requirement that derives its force merely from the arbitrary conventions of one's tribe. Here we have a sort of Euthyphro question: Are the mores good because they prescribe the good, or is the good good because it is prescribed by the mores? To which we may well say: some of each. It all depends, for example, on whether it is driving or murder under discussion.

If there is nothing beyond the individual or collective will, if there is no way to understand how such will could possibly *go wrong*—since there is nothing with which it needs to agree—then what possible sense could there be in our "state of great confusion and uncertainty (some might even say chaos) about what norms or 'universals' ought to govern our practical lives"? Either there is something beyond that decision, relative to which it can be evaluated, or it makes no apparent sense for us to worry over which decision to take or which rule or convention to adopt, since whatever it turns out to be would be right by definition.

V. SECOND ARGUMENT AGAINST SERIOUSNESS: DISAGREEMENT

The second main argument against serious philosophy calls it a monumental failure that has had its chance and deserves no more attention, since it has not "paid off." And just why has it failed, why has it worked so poorly? The fundamental reason to which the rhetoric of failure reduces here is our remarkable lack of agreement both diachronically and synchronically. But why exactly is agreement so

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

important? Is any institution or activity within which we fail to reach substantial agreement one to be rejected and abandoned? Is any institution or activity that yields extensive agreement to be embraced? In that competition, Athens is no match for Sparta, and fanatical cults deserve our respect and support.

Clearly it cannot be agreement per se that is relevant and important. But then what sort of agreement shall it be? One plausible restriction is anathema to the free: that it be agreement whose existence is best explained by the correctness of its object. But really that is nothing special about agreement in particular; it is rather a feature shared by our legitimate faculties and methods generally: including perception and memory, for example. Agreement that *p*, ostensible perception that *p*, ostensible memory that *p*—the acceptability of such profferings is always aided by the explicability of the proffering through the truth of what is proffered. In a fuller treatment of the topic, the variety of threads whose interweaving makes up a living tradition would deserve unraveling. Suffice it here to distinguish rules like those which specify the side of the road one is to use, again, from rules such as the moral and legal injunction against premeditated murder. The former rules are just arbitrary conventions which clearly might have been the reverse with no loss, but the latter rules are more than just conventions to facilitate interaction in keeping with ethnocentric goals. (More on agreement below.)

Serious philosophers believe in a reality independent of being minded by us, though one knowable by us. That reality is said to provide us not only with objects of knowledge, but also with compelling starting points of reason, along with its obligatory methods. And these starting points and methods are supposed to constitute a framework of reason that determines the conditions and extent of rationality, knowledge, goodness, and rightness. Finally, uncovering more and more of this framework is held to be a main goal of philosophy. Thus, for the serious, there are facts of nature and of logic not derived from anyone's thoughts or decisions, and there are moral desiderata with a status and validity not derived merely from the individual or social will (each of which might conceivably have willed quite differently, and each of which might even have willed in a way opposed to the way in which they do in fact will). For example, it is a fact of nature that the earth is round, and the earth might have been round even if there never had evolved or otherwise existed any people or other finite thinkers. As for a particularly simple fact of logic, take the principle of identity that nothing is diverse from itself: even if there never had been finite thinkers at all, still nothing would

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

have been diverse from itself. Finally, inflicting pain just for the sake of doing so, just for its own sake, is something that would have been evil even if we all had converted to some cult that urged the opposite.

Free spirits are radically opposed to the spirit of seriousness. For the free there is absolutely nothing independent of the collective will of the tribe: neither natural facts, nor facts of logic, nor moral desiderata. How, then, can they respond to the intuition that the roundness of the earth and the impossibility of self-diversity are independent facts whose status does not derive from anyone's acceptance or desire? The earth might still have been round even if life had never evolved and there never had existed any finite thinkers at all. And the earth was round and its surface layers contained various distinct substances even before the evolution of life (not to mention thought or consciousness). Otherwise how to explain such evolution? For life and consciousness to have evolved, must there not have been some pre-existing substances, or at the very least must there not have been some pre-existing events or processes not yet involving life or consciousness, namely the evolutionary events or processes that led eventually to life or consciousness? And however imperfectly known by us, those events must have had some definite nature and character, if we are to suppose that they would causally explain how it is that out of them life or consciousness eventually evolved.[11]

VI. WHAT TO MAKE OF OUR INTUITIONS?

Speaking of intuitions, a leading free spirit argues as follows.

> [The neopragmatist's] *only* argument for thinking that these intuitions and vocabularies should be eradicated is that the intellectual tradition to which they belong has not paid off, is more trouble than it is worth, has become an incubus. . . . Dogmatism of intuitions is no worse, or better, than the pragmatist's inability to give noncircular arguments.
>
> This upshot of the confrontation between the pragmatist and the intuitive realist about the status of intuitions can be described either as a conflict of intuitions about the importance of intuitions, or as a preference for one vocabulary over another. The realist will favor the first description, and the pragmatist, the second. It does not matter which

[11] Lucid defenses of strong realism are offered by William Alston in "Yes, Virginia, There is a Real World," *Proceedings and Addresses of the American Philosophical Association*, LI (1979): 779–808; and by Alvin Plantinga in "How to be an Anti-Realist," *Proceedings and Addresses of the American Philosophical Association*, LVI (1982): 47–70. Plantinga also offers a way to accommodate some degree of antirealism by appeal to theology. And a further interesting *via media* is due to Hilary Putnam, *Reason, Truth and History* (New York: Cambridge, 1981).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

> description one uses, as long as it is clear that *the issue is one about whether philosophy should try to find natural starting-points which are distinct from cultural traditions, or whether all philosophy should do is compare and contrast cultural traditions.* . . . The intuitive realist thinks that there is such a thing as Philosophical truth because he thinks that, deep down beneath all the texts, there is something which is not just one more text but that to which various texts are trying to be "adequate." The pragmatist does not think that there is anything like that. He does not even think that there is anything isolable as "the purposes which we construct vocabularies and cultures to fulfill" against which to test vocabularies and cultures. But he does think that in the process of playing vocabularies and cultures off against each other, we produce new and better ways of talking and acting—not better by reference to a previously known standard, but just better in the sense that they come to *seem* clearly better than their predecessors.[12]

"What really needs debate between the pragmatist and the intuitive realist," says Rorty, "is *not* whether we have intuitions to the effect that 'truth is more than assertibility' or 'there is more to pains than brain-states' or 'there is a clash between modern physics and our sense of moral responsibility.' Of *course* we have such intuitions. How could we escape having them? We have been educated within an intellectual tradition built around such claims. . . . So the pragmatist's quarrel with the intuitive realist should be about the *status* of intuitions—about their *right* to be respected—as opposed to how particular intuitions might be 'synthesized' or 'explained away'. To treat his opponent properly, the pragmatist must begin by admitting that the realistic intuitions in question are as deep and compelling as the realist says they are. But he should then try to change the subject by asking, 'And what should we *do* about such intuitions—extirpate them, or find a vocabulary which does justice to them?' " (XXIX).

But that is not really to change the subject; what one should do is always part of the subject. And there is indeed a beautifully simple and conclusive way to end the problems of life: to bite the bullet and put it in one's brain. Short of that one can opt for the simple life of a monk, or a flower child, or a soldier, or any of a thousand others. Many are the ways to deaden oneself to the tortures of philosophy, but the more drastic deaden one also to its consolations. And, if one stops short of the final solution, then how far short? And on what basis? Again and again we are put through the litany of tradition, solidarity, community, forms of life, etc. Only two of the greatest will have none of this: one appeals directly to God through subjective

[12] Rorty, *Consequences*, p. xxxvii.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

truth, the other rejects the "herd" for the will to power of the superman.

Of course one can stop thinking about philosophy. How many ever really start? One can also stop thinking about history and geography and astronomy. One can refuse to look through telescopes, or to inspect maps, or to pore over documents. Is *that* all worth the trouble? And, if we are steadfast, we shall have little to say and little to worry about concerning the heavens, or the oceans, or the historical past. And, to lighten our burden of ignorance, we might even deny that there really are any such things, or better yet purge our language of all pertinent vocabulary. One might even just take a personal vow, and let the community do as it will. Again, why the obsession with solidarity, why the fetish of community agreement? Indeed, why believe in a community at all? Once we reject belief in an independent earth, and belief in an independent law of identity, and belief in the independent evil of pure malice, who can support a belief in others independent of oneself? What then am I to believe? That all the others are dependent on myself and my thought? Maybe the superman can manage that, or those who are willful enough to derive the very divinity from their subjective truth. The rest of us are not that egocentric. So how then are we to think of this community whose form of life is supposed the bedrock of our being, solidarity with which is exalted as the true meaning of our lives? Is each of its members dependent on the thought and vocabulary of all the others? And are all the others dependent on each?

That is, of course, a fatal flaw of any facile idealism. But, if we urge it against neopragmatism, we shall be missing the point. For the point now is not that the earth is dependent on us and our form of life. When realism is rejected by the neopragmatist, it is not rejected in favor of idealism or neutral monism or any other traditional "ism."[13]

[13] Other free spirits may stay closer to the tradition, however, at least in the way suggested by Derrida when he writes that the "step 'outside philosophy' is much more difficult to conceive than is generally imagined by those who think they made it long ago, with cavalier ease, and who in general are swallowed up in metaphysics in the entire body of discourse which they claim to have disengaged from it" (p. 288). Compare also: "the passage beyond philosophy does not consist in turning the page of philosophy (which usually amounts to philosophizing badly), but in continuing to read philosophers *in a certain way*" (p. 284). (Apparently, though "there has never been anything but writing," there is now to be no more writing, at least not by those who move beyond philosophy, since they shall stick to *reading in a certain way*.) An example of the step "outside philosophy" into the jaws of metaphysics may be the pronouncement that the subject itself "is not at the center . . . but on the periphery, with no fixed identity, forever decentered, *defined* by the states through which it passes" [Gilles Deleuze and Félix Guattari, *Anti-Oedipus:*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

VII. NEOPRAGMATISM

Whatever some cohorts may be up to, the neopragmatist rejects realism not in favor of some alternative metaphysical "ism," but only because the terms of the debate are rejected. So we must not press our neopragmatist to explain the ontological status of his community. The very notion of ontological status is now rejected, and cannot be brought against him without begging the question. There is the earth and there is also the community, and it is not true to say that the earth is independent of the community and its vocabulary and its form of life, nor that the community's members are independent of each other. Nor, of course, is there any dependence either, in *any* direction. Instead, the very notion of dependence is rejected along with its useless and outmoded way of thought. Naturally, that does not mean we stop believing in the earth, or denying self-diversity, or rejecting pure malice as unqualifiedly evil.[14] We just stop accepting the notion of metaphysical necessity, and, once necessity is rejected, metaphysical modality is rejected, and with metaphysical modality goes metaphysical dependence as well.[15]

But again we must return to our question: Is not *any* field of thought rejectable by rejecting the intuitions proper to that field? Take away the data and you clear away the theory. Take away the telescope, and even Galileo is helpless. Take away the documents,

---

Capitalism and Schizophrenia (Minneapolis: University of Minnesota Press, 1983), p. 20; my emphasis.] Shades of Hume's bundle theory, through the neutral monism of James and Russell, on to Derek Parfit's process selves. *That* debate is about as serious as can be.

[14] But, if we do believe in the earth and its roundness, then is there not after all, in our view, something distinct from all the texts, namely the earth and its being round (approximately), such that, if a text says that the earth is flat, then because the earth is round (which is no text) that text is inadequate (*contra* Derrida and Rorty above, and textualists generally)? And, if we do reject pure malice as unqualifiedly evil, and, if we approve of health and physical wellbeing as so far unqualifiedly good, are these not after all purposes relative to which we may evaluate anything, including the construction of vocabularies or cultures? See again Derrida and Rorty, above. And compare Arthur Danto's "Philosophy as/and/of Literature," *Proceedings of the American Philosophical Association*, LVIII, 1 (1984): 5–20: "it is altogether compatible with being united through a network of reciprocal effects that a literary work should refer, as it were extratextually, though the reference may be complicated as much by intra as by intertextual references. The *Prelude* and *Finale* of *Middlemarch* refer reciprocally, as well as to the novel they frame, and both refer or allude to Saint Theresa, herself not a text save in so wide a sense as to make . . . [a textualist theory that would admit her as just a plain text, a] theory timid and disappointing" (p. 13).

[15] By *metaphysical modality* I mean the notions of necessity, possibility, etc.) studied and systematized by the discipline of modal logic. The notions of metaphysical dependence and ontological status are closely related to the metaphysical modalities, if not even definable in their terms.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and even Gibbon is silenced. Take away modal intuitions, and even Leibniz is powerless.

There is anyhow, finally, the whole question of how to make sense of improving our practice, whether moral or intellectual. If the will, individual or collective, is the only "foundation" of practice; if, better, there is no such thing as a foundation of practice, if there is just what we do, period, then in what conceivable sense can we be *wrong* at any point in any of our practice? Even if we contradict ourselves, why not affect a Whitmanesque insouciance, and let it go at that? If our body of beliefs has no room for modality, how can it frame to itself the *possibility* that it is somewhere wrong? And the same goes for moral practice.

Finally, what are the prospects for philosophy if we abandon the spirit of seriousness and adopt the spirit of freedom? The free spirit often remains philosopher enough to make a general proposal. Rorty's neopragmatism, for example, is a philosophy of coping, and of seeing how "things hang together," a notion from Wilfrid Sellars which he repeatedly employs. And here is the sort of thing we are told. First, with regard to coping:

> The pragmatist does not think that there is anything . . . isolable as "the purposes which we construct vocabularies and cultures to fulfill" against which to test vocabularies and cultures. But he does think that in the process of playing vocabularies and cultures off against each other, we produce new and better ways of talking and acting—not better by reference to a previously known standard, but just better in the sense that they come to *seem* clearly better than their predecessors.[16]

As for hanging together:

> Pragmatism denies the possibility of getting beyond the Sellarsian notion of "seeing how things hang together—which, for the bookish intellectual of recent times, means seeing how all the various vocabularies of all the various epochs and cultures hang together . . . In a post-Philosophical culture it would be clear that that is *all* that philosophy can be. It cannot answer questions about the relation of the thought of our time —the descriptions it is using, the vocabularies it employs—to something which is not just some alternative vocabulary.[17]

Finally, in a recent pronouncement, Rorty throws into prominence the importance now assigned to *play:*

> My hope is that now, at the end of the century, we philosophers may be in a position to regain Hegel's sense of cultures and languages as match-

[16] *Consequences,* p. xxxvii.
[17] *Ibid.,* pp. xxxix-xl.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

> ing themselves against past and future cultures and languages rather than against such extra-human forces as God, the moral law, or "the real world." This sense of ourselves as engaged in a process of reweaving our beliefs and desires rather than trying to bring these into conformity with something else lets us reappropriate Schiller's sense of play as the highest possibility for human life.[18]

Play is perhaps the highest possibility for some free spirits, but more serious philosophers could not be expected to agree. And, if to cope is just to be content or satisfied by your lights or by those of your culture, and, if that is our worthiest objective by the philosophy of coping, how can that philosophy stand up to the soft sell of the hidden persuaders, or the hard boot of political tyranny? The philosophy of coping has no way to conceive how we could be driven from the sunlight to a hideously worse piglife in a cave if the herdsmen take care to make us well satisfied: more satisfied and less dissatisfied than we are now in any case, impoverished though our lives would become. Feeling good and, with foreshortened horizons, quite content, we would even be denied the concepts to be *able* to feel discontent in a great variety of dimensions previously accessible to us. Our lives would still be *better,* since they would come to *seem* clearly better.

"The Philosophy of Coping," it might be replied, "has the internal resources to cope with our objection. The hidden persuaders and the tyrannical boot have a very delicate job to do. They cannot just excise historical memory, for this would not leave room for the situation of their victims to seem better than what went before. The victims must be allowed to retain the capacity for comparative historical judgment. But they cannot be allowed too much historical memory, lest they recall enough to recognize their loss. Even supposing a job so delicate to be actually practicable, the copers might still cope by noting the repugnancy of the outcome *from our present perspective.*"

The power of charisma (Hitler, Mussolini, Reverend Jones), the effects of fads and fashions on human preference, the plasticity of taste (the phenomenon of acquired taste, for instance)—anyone impressed by such phenomena will be less sanguine about the scope of unscrupulous power and authority. Suppose now that the greater value of "new and better ways of talking and acting" does derive simply from their coming "to *seem* clearly better than their predeces-

[18] "From Logic to Language to Play," *Proceedings and Addresses of the American Philosophical Association,* LIX (1986): 747–753. (An address at the Guadalajara Congress of the Interamerican Society of Philosophy.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sors." How then am I *now* to choose between receiving *A* or *B* at *t*, if both (i) now I prefer receiving *A* at *t*, but (ii) now I also know that, at *t*, I would prefer receiving *B*? Are (i) and (ii) so much as rationally compatible? We need to hear more from neopragmatism.

As for hanging together, what is that supposed to be? Historically, one associates this, first of all, with that apotheosis of modality which is objective idealism, with its coherence theory of everything. But that will hardly appeal to those who refuse to look above the concrete and contingent. With the rejection of modality we lose our best account of logical implication or entailment, without which we are clueless on how to understand coherence; and, if hanging together is not coherence, then what grasp are we left on what it might be? Is it something one should want to understand or describe at all, much less make the focus of one's lifework as philosopher?

VIII. SUMMARY

Free spirits reject the basis for serious philosophy by rejecting its key concept of necessity, and with it all logical modality. Candor compels the recognition that we do have modal intuitions, but these the free spirit would ignore or eradicate. For they are worthless fragments if detached from the intellectual tradition that gives them a place. And that whole tradition is to be deconstructed or demolished in favor of something more streamlined and functional, something that anyhow will "pay off" better. Exactly why, pray, has the tradition not "paid off"? Two main reasons are offered. *First,* the tradition has yielded only Derrida's "diaspora of institutes and languages, . . . the publications and techniques that follow on each other,"[19] a Babel of journals and societies, schools and movements. *That* is not the way to scientific objectivity and progress. *Second,* the tradition's reason, bound by objective axioms and algorithms, is no longer credible. Phronesis—weighing of reasons, judgment—displaces calculation and deduction as a better understanding of how we arrive at beliefs and decisions responsibly and justifiably. Relevant facts are, of course, still needed along with relevant values as inputs to our use of phronesis, even if such values conflict incommensurably. But these factual and evaluative inputs derive for the free spirit from the tribe and not from any objective reality; they come rather with our Gadamerian "prejudices and traditions." Accordingly, we must dedicate ourselves not to reflecting dead reality but to improving dialogic community, not to objectivity but to solidarity.

[19] *Writing and Difference,* Alan Bass, trans. (Chicago: University Press, 1978), p. 81.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I argue to the contrary. *First,* the dichotomy of objectivity or solidarity is false. We want *both* inquiry *and* community, and often we want each in part for the sake of the other. *Second,* a sane acceptance of phronesis does not mean that abstractly *anything goes,* and that *only* the individual or tribal will can appropriately determine what goes in the here and now. Historicism and relativism are just blatant nonsequiturs from the denial of algorithmic determinacy. *Third,* why the fetish of community agreement? On our best present bet, that commits us to Big Brother and the police state. There are many valuable institutions within which agreement is not on the horizon—not only those of politics, but also those of psychiatry and other branches of medicine, for example—and agreement even in the sciences is a shifting matter of degree. There is institutionalized agreement within the great religions, indeed more than in any science, but not among them. Do we decide conflicts by counting heads? Agreement is not an end in itself, and, even if it were, it would not warrant just any means, not, for example, the lobotomy of hard questions whose depth eludes agreement. *Finally,* we can, of course, close our eyes to the intuitive data of philosophy, or allow Big Brother to brainwash them away. But what would that show? And how can we possibly defend the denial of intuitive data except by leaning on other intuitive data? So far the free spirit leans on the denial of algorithmic calculation and, often, on the adoption of agreement and solidarity as our primordial goal. But this yields the arguments already considered, and we offer now our response in dialogic solidarity.

ERNEST SOSA

Brown University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE WELL-ROUNDED LIFE*

THIS paper discusses a moral problem, but it is not the most important moral problem in the world. It is not nuclear deterrence, abortion, or the relief of starvation. Still, it is a problem which I hope readers find engaging, and on which I hope they have views.

The problem arises within the moral view called perfectionism, so I must first explain what perfectionism is.

I. ARISTOTELIAN PERFECTIONISM

Perfectionism holds that there are certain intrinsically valuable states of human beings, and that each of us should maximize the achievement of these states both in herself and in others. Like utilitarianism, then, it is a species of consequentialism. It thinks the morally right act is the one which produces the most good. But perfectionism differs from utilitarianism about what this good is. Whereas utilitarians identify the good with happiness, perfectionists find it in such states as knowledge, artistic creation (or contemplation), the carrying out of complex and difficult projects, and friendship. They want us to lead the best life, and to make its conditions available to others. But what determines the best life is not pleasure or satisfaction, it is the achievement of certain objective goods. Perfectionism is the morality of Plato, Aristotle, Thomas Aquinas, G. W. Leibniz, G. W. F. Hegel, Karl Marx, Friedrich Nietzsche, F. H. Bradley, and (in part) G. E. Moore and Hastings Rashdall. So it has been prominent in our tradition. But, for several bad reasons, it has been ignored by philosophers since about 1920. It is time for a reconsideration.

Perfectionism might seem less important, if the activities best by its lights were always most satisfying or brought the most happiness. In various forms, this has been held by Aristotle, J. S. Mill, and John Rawls. But the view is surely naive. It is a commonplace that knowledge can be painful, and serious achievement a strain. An artist's life, balancing joy in successful creation with the frequent anguish of failure, can be less contented overall than the life of an idle clubman. And it is clearly less contented than life in *Brave New World*. So it matters that perfectionism finds it better. And, even when good activities are most satisfying, a question of theory remains. Are the

* For their help some years ago, my thanks to R. M. Hare, Gavin Lawrence, Michael Lockwood, Amartya Sen, and Charles Taylor. For more recent suggestions, the same to G. A. Cohen, David Copp, Dennis McKerlie, Christine McKinnon, Derek Parfit, Stephen Perry, Donald Regan, and Terry Teskey.

0022-362X/87/8412/0727$02.00 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

activities good because they are perfect, or because they are pleasant? Here perfectionism says, as Bradley did, that, even if life without pleasure is inconceivable, "what we hold to against every possible modification of Hedonism, is that the standard and test is in higher and lower function, not in more or less pleasure."[1]

The problem I want to discuss can arise in any perfectionism, but it is most usefully discussed with a specific version in view. So let me briefly describe what I consider the most plausible perfectionism, a theory called *Aristotelian perfectionism.*

Like Aristotle's, this theory starts from the idea that, at the deepest level, the intrinsic goods develop properties fundamental to human nature. A little vaguely, they develop properties essential to us qua human organisms. This idea generates three major goods, of which the first is *physical perfection.* This perfection develops the bodily properties we share with other animals, and is present to a low degree in ordinary good health, and more completely in vigorous athletic activity, especially at the highest, perhaps Olympic levels. So the first Aristotelian claim is that a life is better and more completely human, if it is free of disease or injury and marked by some demanding physical feats. The remaining goods—the more important ones, it will emerge—are exercises of rationality, both theoretical and practical. On the theoretical side, the side of our beliefs, we achieve more *theoretical perfection* the more knowledge we have. The more we understand the world, and ourselves, and our place in the world, the better and more choiceworthy our lives. Of course, not all knowledge has equal value. Knowing the co-stars in some 1930s movie is not as important as knowing a fundamental law of the universe, or understanding the workings of a friend's personality. We need a test for the best knowledge, and I suggest it is the most organized or systematic knowledge, with general principles unifying and explaining derived particulars. This is most clearly present when we grasp a whole scientific theory from first principles down to particular explanations. But it is also present in interpersonal understanding and even the craft knowledge of skilled artisans. In any case, the second Aristotelian claim is that a life is better the more it is informed and aware, and the more it has unified its knowledge in one structure. The last good is *practical perfection,* or the exercise of rationality in action. Parallel to knowledge is the successful achievement of one's goals, given a justified belief that this would happen, or, more briefly, nonlucky achievement. This is the basic practical

[1] "Mr. Sidgwick's Hedonism," in *Collected Essays,* vol I (New York: Oxford, 1935), p. 97.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
value, but, again, some achievements are more valuable and exhibit more rationality than others. Just as it is better to know a scientific law than to know about 1930s movies, so it is better to achieve a fundamental reform of one's society than to light a cigarette successfully. The question how goals are ranked is difficult, but I would say the best achievements involve the same organizing structure in a person's intentions as is present in the best knowledge. On this view, there is practical perfection in any life which is organized around a single end, or in which large parts have one end, so many activities serve a long-term goal; in activities that are complex, challenging, and difficult, especially if they require precise, and precisely-ordered decisions; in political leadership and co-operation with others, which spread one's concerns beyond the self; and in friendship, especially when it is expressed in a nuanced emotional responsiveness. These are all practical excellences or exercises of rationality in action, and they make a life better by their presence.

Much more could be said about each of these goods, but I want to discuss something different. In our lives we often face choices between the goods. On a particular afternoon we may be able to read a book on European history or campaign for political office. Here we can increase our theoretical or practical perfection but not both, or not both equally. More generally, we must choose between a life of predominantly theoretical achievement—an intellectual life, if you like—a life devoted to action, and a life that tries to balance the two. In these cases, we have before us two ideals: knowledge and action, or the thinker and the doer. Should we prefer one ideal to the other, and, if so, which one? Or should we try to combine the two? And the point is quite general. Whenever there are distinct or plural perfections, we will face choices between them, both at particular moments and for our lives as a whole. After defining its individual goods, a perfectionist morality must help us make these choices. It must say how our overall perfection arises out of the individual perfections we achieve, or, more technically, how we can compare perfections when they conflict.

So far as I can see, the structure of perfectionism sets no limits here, nor does an Aristotelian view of human nature. Together they yield a list of perfections, but say nothing about how they are weighed. Nothing in the bowels of perfectionism requires one kind of comparison over others, and the strongest perfectionism can use whatever comparative approach seems most attractive.

II. LEXICOGRAPHIC AND CONSTANT COMPARISONS

For many perfectionists this approach is very simple. They say there is one perfection, often theoretical perfection, which we should

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

always seek ahead of others. If there are different duties to pursue different goods, they give one of them absolute or *lexicographic* priority, so its demands always have moral precedence. Aristotle takes this line when he ranks contemplation ahead of any practical perfection. He wants us to "strain every nerve" to develop the best part of us, namely the intellectual part, implying that we should prefer any intellectual activity to the exercise of practical virtue.[2] Aquinas has a similar view about the relation between bodily development and rational perfection in general. Unlike Aristotle, Aquinas thinks bodily development has some value,[3] but this value is "ordained to" spiritual goods and must always yield before them.[4] Both writers make one perfection infinitely better than the rest, so even a tiny amount of it outweighs a great achievement of the others. Only when we can no longer promote the higher excellence should we seek a lesser good.[5]

Common though it is, this lexicographic approach is too extreme to be plausible. Each Aristotelian perfection has independent attraction, and none deserves to be placed so far above or below the others. Physical perfection seems to me less important than the two rational excellences. I have no special argument for this, but the development of our bodily powers seems to have less value than the exercise of our distinctive rationality. So it should count for less. But to make any increase in rationality, no matter how trivial, outweigh the greatest athletic achievements is to go too far. A plausible morality must give each of the Aristotelian goods some serious weight.

The simplest such morality employs *constant trade-offs* between the different perfections. It assigns a fixed, finite weight to each excellence, and this weight is the same for all persons in all situations.

[2] *Nicomachean Ethics,* 1177b26–1178a7; also *Eudemian Ethics,* 1249b6–25.

[3] According to Aquinas, the perfection of our bodily parts is a necessary component in the blessed state that will follow the resurrection of the body (*Summa Theologiae,* I-II, q. 3, a. 3; q. 4, a. 5; also Supp., qq. 82–84), and also characterized the (slightly less perfect) state that preceded the Fall. Arguing against Origen that humans would indeed have reproduced by sexual intercourse had they remained longer in a state of innocence, Aquinas adds that this intercourse would have been vastly more pleasurable than anything we experience now, because of "the greater purity of nature and the greater sensibility of the body" (*ibid.,* I, q. 98, a. 2).

[4] *Ibid.,* II-II, q. 152, a. 2, 3.

[5] A theory with this view need not require us always to seek the highest excellence in ourselves. If it is agent neutral, as suggested above, it can tell us to sacrifice some of our own achievement of this excellence to promote it in others. This seems to be the view of Aquinas. While agreeing with Aristotle that contemplation is in itself better than any practical good (*Summa Theologiae,* II-II, q. 182, a. 1), Aquinas thinks it meritorious to abandon one's own contemplation to promote it in others (*ibid.,* II-II, q. 182, a. 2, q. 184, a. 7; also III, q. 40, a. 1, ad 1). Since "it is better to illumine than merely to shine" (*ibid.,* II-II, q. 188, a. 6), the most praiseworthy lives are devoted, as Aquinas's was, to teaching as well as to study.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

If it makes 2 units of theory equal 3 of practice, it does so uniformly. Not only must I weigh theory and practice in this proportion, so must you and everyone else.[6] Because there is no pre-existing scale on which perfections are measured, this view cannot in the strict sense make one higher. (If the units for theory were halved, the above proportion of 2:3 would become 4:3.) But it can do this informally. If our comparative principles imply that most people should spend more time pursuing one good than another, there is a clear sense in which they make it more important. This is just what we should want for rational *versus* physical excellence.

Constant trade-offs still permit Aristotle's view that theory is better than practice, now read as the nonlexicographic claim that it deserves more (finite) weight. But, even in this weaker form, the view is implausible. Theoretical and practical excellence both realize rationality, and do so in structurally similar ways. Both are products of evolution, and nothing in their character or origins makes one more desirable than the other. Why should rationality in conduct, in how we change the world, count less than rationality in how we form our beliefs? Why should a structure of ends have less value than a similar structure of convictions? Aristotle's arguments for preferring theory are unimpressive. His central argument—that theoretical excellence realizes a separable and divine element in our nature[7]—is of merely historical interest, as is the claim that it concerns the best objects.[8] Remove God and we have no reason to think what theory knows better than what action affects. Nor are his other arguments more persuasive. That contemplation offers pure and continuous pleasure is irrelevant to perfectionism; that it is more self-sufficient is dubious; and that it alone is loved for itself is false.[9] Skill in games is valued not just as a means to winning—many of us care little for that—but because in itself it exercises rational capacities. Just like contemplation, it can be the object of intrinsic appreciation. Without better arguments than Aristotle's, perfectionism should give the rational excellences roughly equal weight, so a world of uniformly good lives devotes roughly equal time to each.

[6] Nothing in these remarks implies that comparisons can realistically be made with such precision, so that they are 2:3 rather than, say, 21:30. Nor does it imply that the individual excellences can be cardinally measured, as the mere possibility of such precision requires. The idea is just that we can deepen our understanding of comparative judgment by asking what form precise trade-offs would have if they were possible.

[7] *Nicomachean Ethics*, 1177b26–1178a8.

[8] *Ibid.*, 1177a21.

[9] For these three arguments, see *Nicomachean Ethics*, 1177a24–b25; and, for a reprise, Aquinas, *Summa Theologiae*, II-II, q. 182, a. 1.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

III. BALANCING

Our initial suggestion, then, is to have constant trade-offs between the excellences, with physical perfection ranked lower and the rational goods roughly equal. This is not positively unattractive, but I think many of us are drawn to something different. We think the best lives contain a certain balance between the perfections, and do not concentrate too much on one. To talk only of Leonardo would pitch the ideal too high, for what I intend is possible in all lives. We all can spread our activities between different goods, aiming at a well-rounded achievement rather than any narrow specialization. Even if our specific achievements are not great, their proportion can mirror that of Renaissance lives, and for many of us this proportion is, other things being equal, a good. At however modest a level, it gives our lives an intrinsically desirable shape.

This *balancing* view, however, excludes constant trade-offs. Imagine that one person has devoted most of her life to politics, while another has worked only at scholarship. Well-roundedness implies that the first will improve her life most by acquiring some knowledge, while the second should become more active. They should prefer what they have neglected, and what they have neglected is different. At the heart of balancing is the idea that a perfection's relative value depends on the relative amount of it one has achieved in the past. Going beyond equal weights, it says that, if one excellence has been achieved more than another, the second is more important. The clearest representation of this is on an indifference graph (Figure 1). Here each curve links points representing combinations of the two goods that make for the same overall perfection, with curves farther from the origin better than ones close in. For simplicity, imagine the axes calibrated so that one unit of each perfection is what an average person can achieve in a fixed time, say one hour:

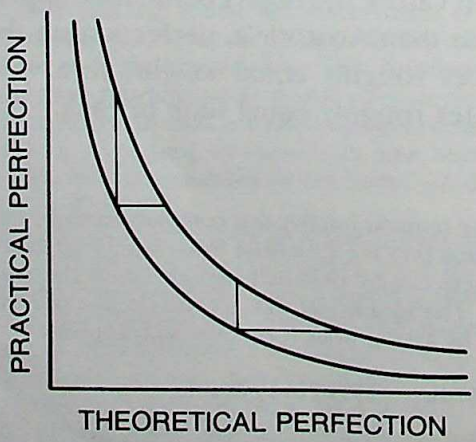


Figure 1

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

In the upper left of the graph, where the politician has achieved more practice than theory, the same overall improvement results from a large increase in action or a small increase in knowledge. So, for her, knowledge is more valuable. In the lower right, by contrast, the scholar makes an identical gain by improving theory greatly or practice a little. Given different past achievements, there are different values now.

Indifference curves are borrowed from economics, where they are used to represent consumer preferences among different commodities. But there is an important difference between the economists' use and our own. On a natural view, the economists' curves reflect the fact that the absolute utilities of commodities diminish as their absolute levels increase. Thus, the reason we prefer an apple to an orange when we have twice as many oranges as apples is that we care less for either fruit the more of it we have. This cannot be the point in perfectionism. Imagine that one person has 10 units each of theoretical and practical perfection, while another has 20. If the value of balanced lives rested on diminishing *absolute* values, the second person's life would be less than twice as good as the first's, which we do not want to say. Our aim is to appreciate Leonardo, not to minimize his feats. For this reason, acceptable balancing principles must look only at a person's *relative* achievements, so ratios between the goods, and not their absolute levels, are what matter.

Given the intuitive appeal of balancing, we should not be surprised that it appears frequently in the perfectionist tradition. Several writers say that an ideal life contains a "harmonious" achievement of different goods. Sir William Hamilton defines perfection as "the full and harmonious development of all our faculties";[10] Wilhelm von Humboldt calls a human's true end "the highest and most harmonious development of his powers to a complete and consistent whole";[11] and Rashdall describes a life where "many distinguishable elements are harmonized and combined."[12] At times the point of these remarks is causal. Humboldt says a person can "increase and diversify the powers with which he works, by harmoniously combining them, instead of looking for a mere variety of objects for their separate exercise."[13] This is sound advice on any view of comparison.

[10] *The Metaphysics of Sir William Hamilton,* F. Bowen, ed. (Boston: John Allyn, 1872), p. 1.

[11] *The Limits of State Action,* J. W. Burrow, trans. (New York: Cambridge, 1969), p. 16.

[12] *The Theory of Good and Evil,* vol I (London: Oxford, 1907), pp. 220/1. Elsewhere Rashdall says that "no amount of one good can compensate for the absence or deficiency of the other" (vol. II, p. 39), an even clearer balancing claim.

[13] *The Limits of State Action,* p. 16.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

If some activities, or ways of organizing activities, enable us to achieve two excellences at once, any perfectionism should find them good. But the talk of harmony goes beyond this, and suggests that some proportion between goods is violated if we concentrate too much on one. The same view is present in the polemics of Marx and Nietzsche against specialization and the division of labor. Marx famously wants to be able "to do one thing today and another tomorrow, to hunt in the morning, fish in the afternoon, rear cattle in the evening, criticize after dinner, just as I have a mind, without becoming hunter, fisherman, cowherd, or critic";[14] while Nietzsche heaps scorn on the "nook-dwellers" and "fragments of humanity" he finds among European intellectuals.[15] Against "modern ideas" that would "banish everybody into a corner and 'speciality'," he insists that man's greatness lies in his "range and multiplicity, in his wholeness in manifoldness."[16]

The exact impact of a balancing view depends on facts about the lives we have available—facts we discuss below (sec. IV). But it also depends on the goods' weighting. Like constant comparison, balancing can make one perfection higher if it tells us to spend more time pursuing it. (This will be reflected in curves that tilt toward one axis, rather than lie symmetrically as in Figure 1.) We should take this line with rational and physical perfection. Then our theory will say that, although it is important for a life to contain some bodily goods, there should usually be less time spent on them than on rationality.

Balancing has several indirect consequences, one of which is to reduce the importance of an excited chase after goods. On any view, the best lives strive vigorously after excellence. But, with balancing, this does not guarantee a close approach to overall perfection. There are many lives of all-out activity that fail, through lack of balance, to equal the value of lazier lives whose proportion is right. Energy matters, but, given balancing, its level can be less important than the direction in which it is expended. At the same time, balancing increases the importance of self-knowledge. With constant trade-offs,

[14] Marx and Friedrich Engels, *The German Ideology,* in D. McLellan, ed. *Karl Marx: Selected Writings* (New York: Oxford, 1977), p. 169; see also Marx, *Grundrisse,* M. Nicolaus, trans. (Harmondsworth: Penguin, 1973), pp. 325, 487/8.

[15] *Thus Spake Zarathustra,* Second Part, "On Redemption," in W. Kaufmann, ed., *The Portable Nietzsche* (Harmondsworth: Penguin, 1976), p. 250. Also *Beyond Good and Evil,* W. Kaufmann, trans. (New York: Vintage, 1966), sec. 204, 205, 212, 257; *On the Genealogy of Morals,* W. Kaufmann and R. J. Hollingdale, trans. (New York: Vintage, 1969), Third Essay, sec. 23; *Twilight of the Idols,* "What the Germans Lack," sec. 3, in *The Portable Nietzsche,* p. 508; *The Antichrist,* sec. 57, *ibid.,* p. 647; and *The Will to Power,* W. Kaufmann, trans. (New York: Vintage, 1968), sec. 390, 881.

[16] *Beyond Good and Evil,* sec. 212.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

a person can compare without knowledge of her past. Since the relative weights of the perfections are always the same, she can choose correctly without knowing how she chose before. Balancing, however, presupposes an awareness of one's history. As a realization of theory, this awareness is intrinsically good, and, by contributing to a life that is unified across time, it also advances practice. Here we see how, given balancing, self-knowledge is also a condition for correct comparative judgment.

Alongside its consequences for the general shape of our lives, balancing affects our choice of specific activities. As Humboldt points out, one way to achieve "harmonious development" is to choose projects that promote several excellences at once. These projects are recommended by any perfectionism, but even more by one with balancing. Scientific research passes this test, for it combines the goods of knowledge and action. A scientific researcher sets herself a goal—discovering a new law, or uncovering the motion of continents—and uses skill and ingenuity to achieve it. Her activity applies and creates valuable knowledge, but it also involves subtle strategic thinking. Something similar can happen in athletics. Team sports can require players to grasp and solve strategic problems at the same time as they exercise their bodies. Here the mix of rational with physical goods is what balancing finds attractive, and what suits these sports to a well-rounded life.

Given this last point, a life could count as well-rounded, in the sense defined up to now, even though it concentrated on just one activity. A life devoted entirely to scientific research could be balanced if its inquiries developed practice as much as theory, and a narrowly hockey-playing life could be rounded if the sport's intellectual demands match those on the body. This may seem counterintuitive, and false to the original ideal of well-roundedness. It may seem far from honor to Leonardo to think his virtues attainable in a life with just one realm of achievement. The result is counterintuitive, but we can avoid it by extending the balancing view from different perfections to different realizations of one perfection. Then it will, other things being equal, be better to have knowledge of European history, astrophysics, and one's friends' characters than to concentrate understanding in a single area, and better to carry out complex plans of different types than to have just one kind of practical achievement. This extension of balancing is natural, and I will assume it in what follows. I will assume that we are to seek variety, not just among different excellences, but among different aspects of one excellence; so a fully rounded life is developed, knowing, and active in several areas. Unfortunately, the extension complicates our view, and makes it harder to represent graphically. So I will often continue

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to speak as if we are to seek balance primarily among our three most general perfections.

IV. DILETTANTISM AND CONCENTRATION

Balancing says that, other things being equal, we should prefer the excellences we have achieved less in the past. But, to apply this advice, we must know when other things are equal, or, more precisely, what other effects a search for balance will have. Does diversity of action undermine the specific goods, so a well-rounded life is one of uniform mediocrity? If so, balance, while attractive in theory, is not practically recommended. Or does diversity enhance the perfections, enriching them from different sources? Then balance is doubly valuable.

To sharpen these questions, let us return to our indifference graph and add an *achievement line*. The points on this line represent the greatest combinations of perfection a person can achieve given her natural abilities, her resources, the time in her life, and whatever else affects her excellence. The area under and including the line represents the total lives available to her, and, in comparing perfections, she in effect chooses a point in this area. If her achievement line is straight and symmetric to the axes (Figure 2), it touches the outermost indifference curve at a single central point, which we can call its *ideal-life point*. Given this line, she leads her best life by giving equal time to both goods.

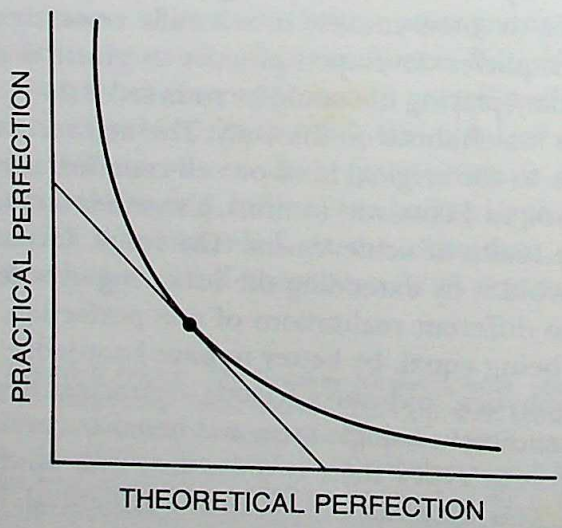


Figure 2

Real-world achievement lines, however, are hardly straight, and their shape can affect the impact of balancing. Imagine, first, a line

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that is convex to the origin, so its shape mirrors that of the indifference curves [line (i) in Figure 3]:

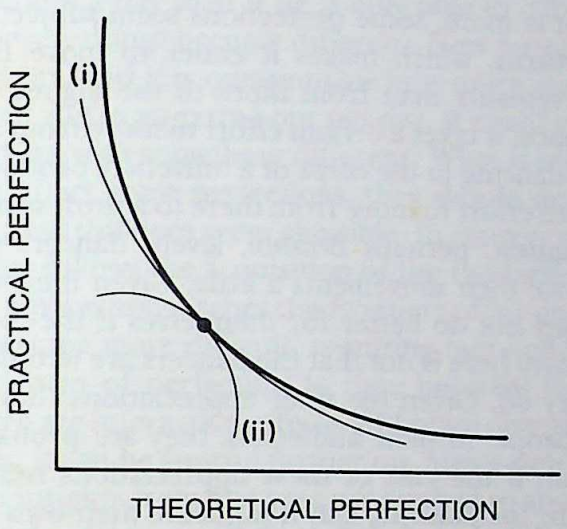


Figure 3

With this line, balance is not such a virtue, nor specialization such a crime. Even very concentrated lives come close to the outermost indifference curve, and, with sufficient convexity, could touch it. By contrast, a line that is concave to the origin [line (ii) in Figure 3] makes balance more important. Concentrated lives fall much shorter of the ideal, and are even more seriously wrong.

Achievement lines belong to individuals, and it would be foolish to think that one shape fits all. Nevertheless, there may be general factors that affect the shape of all lines and determine the general import of balancing. What could they be?

There seem, unfortunately, to be two opposing factors, one supporting convexity and one its opposite. In favor of the convex line, and against balance, is what we can call *dilettante's disadvantage*. People often find that, the more time they invest in an activity, the more they gain from further attention to it. A professional historian can learn more from an hour with a history book than can an athlete, who lacks the background to read it critically. At the same time, the athlete can achieve more physical perfection in an hour's exercise. He can run farther, or lift greater weights. In both cases, the past devotion pays dividends now. It is often said that people who spread themselves across many activities will not achieve much in any. If this is correct, a well-rounded life will be low in all perfections, and, to reflect this, most achievement lines will be convex.

On the other side are the *costs of concentration*. People who devote themselves to one activity can easily become fatigued or even

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

bored. They can miss the invigorating effect of variety, as well as the chance to enrich themselves in one area with experiences gained doing something different. Their very narrowness can produce stultification. What is more, some perfections seem subject to a law of diminishing returns, which makes it easier to move from low to middling achievements than from there to the highest heights. In ballet, for instance, it takes a certain effort to move from a beginner's level to that of dancing in the *corps* of a university production. But it takes much more effort to move from there to a professional production. At the highest, perhaps Bolshoi, levels, dancers practice for hours to improve their movements a little. Given their tremendous costs, might they not do better for themselves if they sought some variety? The worry here is not that the dancers are wrong to practice as much as they do. Given the finer appreciations they permit the thousands of people in their audiences, they are probably, on balance, right. But, if the cost of these appreciations is such narrow concentration by the dancers, may it not leave their own lives impoverished?

These opposing factors are both plausible, and readers may disagree about their relative weights. Some may think the costs of concentration more important, and continue to value well-roundedness even when all is considered. Others may emphasize the disadvantage. But there is a third possibility. We may be able to acknowledge both factors and unite them in one view, if we look more closely at where their force is greatest. Although they pull in different directions, the factors seem to pull most strongly in different places, and may therefore not be directly contradictory.

Dilettante's disadvantage, to start, seems strongest at low levels of perfection and weaker high up. Many arenas of perfection are such that large investments of time are needed before any value results. In a cognitive domain, there are basic concepts and principles to grasp; in a practice, there are fundamental skills, and without them a person's activity counts little. Unless he does enough to acquire the fundamentals, his achievement is minimal. (Consider learning a language: without a basic grasp of the grammar, individual bits of vocabulary mean little.) Once these fundamentals are present, however, new possibilities open. A person can use the basic concepts and skills to acquire others, and, for a time, his progress is rapid. Because of this, diversity is most damaging when it prevents a person from acquiring any fundamentals. But it is less so farther along. Someone with the elements of several perfections can make real gains in them all, and need not be distracted by variety. Where a beginner must concentrate, he can alternate among activities, returning to each with his mastery secure.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

The costs of concentration, by contrast, seem greatest at high

Digitized by Arya Samaj Foundation Chennai and eGangotri

levels. Someone who devotes an hour a day to history is in little danger of fatigue, and can easily add an extra hour. Doing so may even increase his scholarly commitment. He may have more energy given two hours a day than if he is only able to dabble. After eight hours, however, things become different. Now boredom and fatigue are real factors, and it is questionable how much someone studying that hard can add in an extra hour per day. It might even improve his history to relax with something different. What is more, if diminishing returns affect some perfections, they should do so primarily at high levels. And this does seem plausible. In many domains, the rapid progress that follows the acquisition of the fundamentals eventually slows. As a person approaches the frontiers of his discipline, further advances become more difficult, requiring more effort and application. The return of perfection to time becomes less, not more.[17] Finally, there are interaction effects. Although diversity is distracting at low levels, it can be fruitful farther on. Major new advances often occur when insights from one area are applied to another, and those who seek such advances will do well to broaden their experience.

These arguments suggest a unified picture: dilettante's disadvantage is strongest at low levels, and the costs of concentration greatest higher up. The bad effects of variety are most significant when one's achievements are low, and the costs of specializing worst when they are high. And this, in turn, suggests that many people's achievement lines may be M-shaped, as in Figure 4:

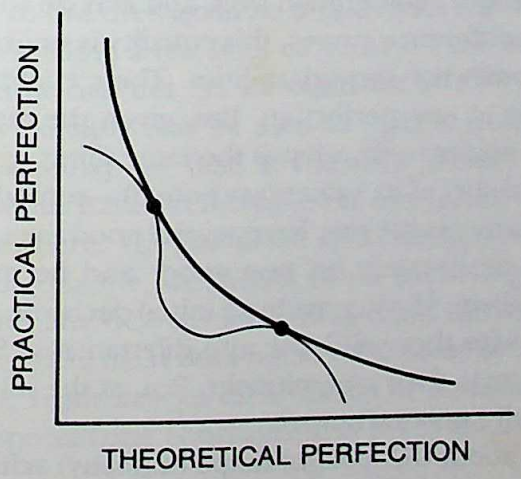


Figure 4

[17] It is this possibility—different returns of perfection at different levels, as well as to people with different talents—that forces us to equate one unit of perfection with what an *average* person achieves in an hour (sec. III). With cardinal measures of the individual excellences, we can compare goods in one area, e.g., Leonardo's paintings with a beginner's. But we need to normalize our measures, and, for this, look to an average person's achievements in a fixed time.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

In the upper left and lower right of Figure 4, where the achievement of one good is quite high, the costs of concentration work strongly on that good to make the line concave. In the center, however, where both perfections are low, the disadvantage dominates and the line is convex. If the M-shape characterizes a person's options, she has two better lives available, each concentrating to some extent on one good. She must be neither a pure specialist nor a pure all-rounder, but give a moderate preponderance to one perfection. To give more than that preponderance is to fall afoul of the costs, which can be serious. But it is also bad to aim at too much balance. Someone who does that will slide into the central trough of her achievement line, and onto a lower curve.

In the configuration of Figure 4, the M-shape has most importance, and the upshot would not be very different with constant trade-offs. But balancing still plays a role. For one thing, it makes the ideal life slightly more balanced, with its point on the inside slope of a hump rather than right on top. So a person's goal is slightly more proportion. Balancing also affects the badness of less than ideal lives. Because the indifference curves dip into the central trough, a dilettantish life is, while not best, less bad than if the curves were straight. And an overly-specialized life is worse. It falls much shorter of the outermost curve, and is more seriously lacking. Even with M-lines, balancing affects deliberation, pushing us toward proportion and away from monomania.

When a person has two ideal-life points, as in Figure 4, she can choose between two equally good lives, and may do so as she pleases. With precise indifference curves, this equality is unlikely, since most people have somewhat skewed abilities. Their talents incline them somewhat more to one perfection. But, given the many curves in a realistic perfectionism—or, what is the same thing, given the merely partial comparability of its values (see note 6)—something like it may be common. Many people may have several good lives available, each concentrating moderately on one good, and be morally free to choose among them. Having made an initial decision, they must stick to it, for otherwise they will lapse into dilettantism. So past a point they must be firm in their commitment. But, at the start, their choice of lives is morally unconstrained.

If I am right about the (rough) shape of (many) achievement lines, they answer the most common objection against balancing, namely that it promotes mediocrity. Rashdall urges this objection against a "self-realization" theory that identifies good with "an equal, all-round development of one's whole nature":

> Up to a certain point the man who is a mere specialist will be a bad specialist, but that point is soon reached. Charles Darwin found that the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

> cultivation of reasoning power and observation had extinguished his once keen imagination and sensibility. And yet who would wish—whether in the interests of the world or in the interests of what was best in Charles Darwin's own nature—that his work should have been spoiled in order that one of the three hours which was the maximum working day his health allowed should have been absorbed by politics or philanthropy? Who would decide that the origin of species should have been undiscovered, in order that the man who might have discovered it should retain the power of enjoying Wordsworth? This notion of an equal, all-round, "harmonious" development is thus a sheer impossibility, excluded by the very constitution of human nature, and incompatible with the welfare of society. And, in so far as some approximation to such an ideal of life is possible, it involves a very apotheosis of mediocrity, ineffectiveness, dilettantism.[18]

This entertaining passage rather exaggerates its history. Darwin "discovered" the theory of evolution some twenty years before publishing it in *The Origin of Species,* and continued scientific work for another twenty years after *The Origin of Species.* So a realistic picture of his life without an hour's biology a day will have it lack, not the discoveries that made him famous, but perhaps the works on *Insectivorous Plants* and *The Effects of Cross- and Self-Fertilization* which he published late in his career. And is it wrong to say that his life would have been more complete had he left the work of these books to others, and done more outside biology? I do not think so, and neither, it appears, did Darwin. In a memoir published by his son in 1929, Darwin is reported to have told one of his daughters that "if he had his life to live over again he would make it a rule to let no day go by without reading a few lines of poetry. Then he quietly added that he wished he had not 'let his mind rot so'."[19]

The Darwin example may be used to make a stronger claim than Rashdall's. Some may say that, if Darwin's greatest talent was for science, he should have preferred even tiny achievements there to large ones elsewhere. He was above all a biologist, they may argue, and, in assessing his life, we should count his biological perfections above all others. The view advanced here is not the lexicographic one discussed above (sec. II). It does not say that some activities are better for all humans, regardless of their talents or situation. But the view still supports specializing conclusions. It holds that, although differ-

[18] *The Theory of Good and Evil,* vol. II, p. 64. How this argument is supposed to mesh with Rashdall's own remarks about "harmonizing" different elements (see above, note 12) I cannot say.

[19] Major Leonard Darwin, "Memories of Down House," *The Nineteenth Century,* CVI (1929): 119–20, quoted in William Irvine, *Apes, Angels, and Victorians* (New York: Time, 1963), p. 247.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ent perfections are in themselves equally good, the value of a person's life depends primarily on what he achieves in his single best one. If his talents incline him more to one good than to others, he should prefer any gain there to improvements in its rivals.

I have little to say about this *pure specializing* view. It directly counters the position I have been developing, and violates the intuitions I most want to capture. Although it is intelligible, I do not think it really survives the application to serious judgments. When we think seriously about lives, the compelling ideal is surely of well-roundedness. And we can retain this ideal if the most common objection to it, namely Rashdall's objection, is answered by M-lines. These lines make too much balance a bad thing; and views that value well-roundedness in the abstract can agree that, if appreciating Wordsworth had prevented Darwin's main discoveries, it would have been wrong. At the same time, they can say that, of the broadly specialist lives available to Darwin, the best contained at least some other goods and were not solely devoted to biology.

I think the combination of balancing and M-lines yields a perfectionism in harmony with many intuitions about specialization. On the one hand, it makes well-roundedness an ideal, and favors lives with the greatest balance consistent with serious achievement. On the other, it reflects our common suspicion of dilettantes. The combination may also help historically. When classical writers discuss theory and practice, they often compare, not knowledge and action as such, but the value of whole lives devoted to them. They do not ask whether understanding is better than political rule (for many the highest practical activity), but whether a philosophical or contemplative life is better than a political life.[20] This makes sense on the view we have described. It follows from balancing that we cannot compare two excellences directly. Since a unit of theory that would mean a great deal to a politician will add little to a scholar, we cannot assess it apart from the life in which it figures. The smallest unit we can compare is the complete human life. And, if this explains the focus on lives, the restriction to specialist lives may follow from M-lines. If we want to compare the best lives commonly available, as most perfectionists do, the disadvantage tells us to ignore any that divide their time equally. We can consider contemplative and political lives, but not ones that aim at both goods. One perfectionist who discusses these lives is Plato, in his account of the philosopher-kings.

[20] Aristotle, *Nicomachean Ethics*, 1095b14–1096a10, 1177a11–1179a33; *Eudemian Ethics*, 1215a26–1215b14, 1216a27–37. See also Aquinas, *Summa Theologiae*, II-II, q. 182.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

But what Plato says is that, precisely because they have to combine such different abilities, these kings will be hard to find.[21]

A final point is this. M-lines are plausible because dilettante's disadvantage is strongest at low levels and the costs of concentration greatest higher up. But, if this applies along one achievement line, it should also hold for a series of lines representing the options available to different people. Those with limited abilities or a short lifespan should find that dilettante's disadvantage works powerfully on their options, creating a deep central trough in their line. For those with more talent and time, however, this factor should be less important. As their achievement lines move farther from the origin, the troughs should become shallower, and the line should eventually approximate the simple concavity that would obtain given only costs of concentration (Figure 5):

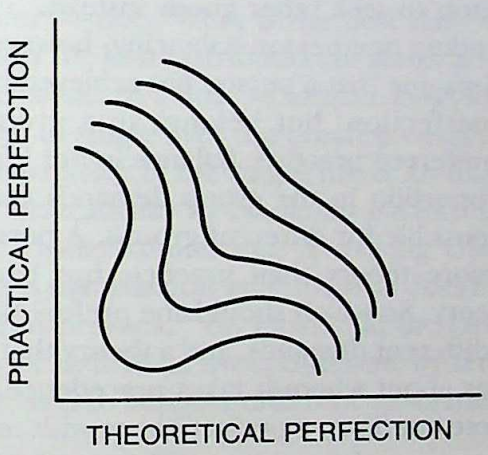


Figure 5

This too matches intuition. For it implies that, although exceptional individuals like Marcus Aurelius and Leonardo can aim at well-rounded lives, those with more limited gifts must content themselves with being some kind of specialist.

V. MANY-PERSON BALANCING

So far we have applied balancing within single lives, and made the relative weight of a perfection depend on the relative degree to which a person herself has achieved it. But it is also possible to balance across lives. We can define an ideal proportion for a collective, and say a group does better when its members achieve a variety of goods than when they all specialize alike. Just as we value well-

[21] *Republic*, 503c.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

roundedness within lives, so we can value it across them, and prefer groups whose members exercise a diversity of talents.

For some perfectionists, this many-person balancing is a simple extension of the one-person approach. In his account of "harmonious development," Humboldt says that what occurs in one person "by the union of the past and future with the present, is produced in society by the mutual co-operation of its members."[22] An individual can only achieve some excellences open to the species, but by entering into a "social union" with others she can overcome this limitation and "participate in the rich collective resources of all the others" (*ibid.*). The intuitive idea here is that, if she belongs to a group whose members achieve all the goods, she can participate through them in a well-roundedness denied her on her own. But the idea again rests on balancing. A group can only offer this participation if some members' achieving one perfection permits others to neglect it, and gives them a reason to seek other goods instead.

Far from extending one-person balancing, however, this view can conflict with it. Imagine that a person has achieved more theoretical than practical perfection, but belongs to a group whose other members have preferred practice. Balance in her life requires more action, while proportion in the group demands the opposite. The same conflict is possible for different groups. A person's family may have achieved more theory than practice, but her country more practice than theory. So which should she prefer? Balance in different units makes different demands, and a theory that extends it faces difficult questions about where it takes precedence.

Because of these conflicts, a perfectionism with many-person balancing will be very complex, with a different overall goal for each collection where proportion is valued. This is not a decisive objection, but I think the strongest perfectionism should confine its balancing to single lives. Let me explain why.

We should agree, first, that the intuitive appeal of balancing is greatest in single lives. There an ideal of proportion strikes us immediately, whereas for larger groups it is less obvious. There is a reason for this. The degree to which balancing is plausible for a collection depends on the degree to which the collection forms a unity. The states making up a single human life are connected in especially intimate ways, and form a tighter unity than any collection across lives. This makes it especially natural to consider them together, and to seek proportion among them. To confirm this note that, when we do balance in groups, it is only the most unified that we consider. We

[22] *The Limits of State Action*, p. 17.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

may praise a family if its members achieve different excellences, or admire a civilization like Greece for excelling in such diverse areas as statecraft, poetry, warfare, and philosophy. But, if we consider accidental collections like the group of people born on a Tuesday, or with names containing 7 letters, the idea does not suggest itself. And is does so even less if a group has no internal connections. Imagine a universe containing two civilizations that live in galaxies light years apart and have no knowledge of each other. Surely it is not plausible to claim that, if one civilization is realizing theory more than practice, this gives the other grounds (which it cannot know) to concentrate instead on practice.[23]

If this is right, decisions about balancing reflect views about where significant unities lie, or about what collections form significant units. Now, perfectionism has always made the complete life its central moral unit. Its first question has always been, "What is the best whole human life?"—not "What is the best state of a person at a time?" or "What is the best distribution of states across persons?" It would reflect this concern a little to balance everywhere, but to give the consequences in single lives the greatest weight. But it is surely simplest to stick to lives. If the single life is an important unit, we signal this most dramatically by confining balance to it.

And there is a wider connection. A striking claim of Rawls's *A Theory of Justice* is that classical utilitarianism does not take seriously the separateness of persons,[24] or, looking at the same fact from the inside, the unity of individual lives. Different writers have used this claim differently, and to support different moralities;[25] but, for me,

[23] In a discussion that has influenced mine, Brian Barry suggests a balancing treatment of two social values that he calls equity and efficiency [*Political Argument* (London: Routledge & Kegan Paul, 1965), pp. 3–8; see also John Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard, 1971), pp. 37–40]. The proposal, however, is only plausible for something as integrated as one society. No one would claim that, if another nation is achieving more equity than efficiency, ours must seek more efficiency to redress the balance.

[24] *A Theory of Justice*, pp. 29, 187.

[25] In Rawls, separateness seems to support a methodology, namely the hypothetical contract, and only indirectly any egalitarianism chosen through it. But, more recently, Thomas Nagel and Samuel Scheffler use separateness to argue for equality directly ["Equality," in *Mortal Questions* (New York: Cambridge, 1979), pp. 106–127; and *The Rejection of Consequentialism* (New York: Oxford, 1982), ch. 3]. And Robert Nozick claims that separateness supports libertarianism, which notoriously ignores equality [*Anarchy, State, and Utopia* (New York: Basic Books, 1974), p. 33]; John Taurek cites it in defending a view that is insensitive to numbers ["Should the Numbers Count?" *Philosophy and Public Affairs*, VII (1976–1977): 293–316]; and Derek Parfit thinks it may support rational egoism [*Reasons and Persons* (New York: Oxford, 1984), pp. 139, 521]. This array of arguments, all, it seems to me, equally plausible (or implausible), makes me doubt whether Rawls's claim has more force than that suggested for it in the text.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

its core is this. The states morality deals with come divided into lives, and have more intimate relations within these lives than across them. This is an important metaphysical fact, and should be reflected in moral thought to the following extent. It should be impossible to apply a moral view conclusively without knowing how the states it deems important divide among persons. If all we know about an act is that it will produce such and such states, without knowing which go into which lives, we should not be able to judge the act decisively. This simple condition is not satisfied by classical utilitarianism, which can evaluate so long as it knows the total happiness an act produces. It does not take seriously the differences between persons, because it does not require us to know about these differences before we judge conduct.

Now, this reading of separateness also supports one-person balancing. We have already noted that this balancing prevents us from comparing excellences directly (sec. IV). An item of knowledge that would be good for a politician will add little to a scholar, and we cannot assess it apart from the life in which it figures. Perfectionist judgments presuppose comparison, and, with one-person balancing, this requires assigning goods to lives. The same is not true of many-person balancing, which can compare once it knows the total in a group. So, on this score, it is no better than classical utilitarianism. It would, again, meet Rawls's point a little to give one-person balance more weight than balance across lives, so it is only somewhat more important. But it is more dramatic to stick to lives. This restricts us to the balancing with the most intuitive appeal, and also acknowledges separateness in the strongest possible way.

The separateness of persons is important, but it does not require as radical a response as some writers suggest. To acknowledge the metaphysical facts, we need not abandon goal-based morality, as Nozick argues, or even summation across persons, as Rawls believes. We can retain a basically consequentialist structure so long as our goods are perfectionist, and, in weighing them, we consider only relative achievements within single lives.

THOMAS HURKA

University of Calgary

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

Decisions and Revisions: Philosophical Essays on Knowledge and Value. ISAAC LEVI. Cambridge: Cambridge University Press, 1984. xvii, 293 p. $34.50.

This volume reprints (with some minor alterations) seventeen of Isaac Levi's philosophical papers. A preface presents the main lines of the historical evolution of Levi's philosophical ideas and also explains the interconnections between the selected papers; it concludes with the bibliography of Isaac Levi. The volume forms an excellent complement to his two books, *Gambling with Truth* (1967) and *The Enterprise of Knowledge* (1980), but it can also function as a relatively self-contained presentation of Levi's thinking. The volume contains the most important of Levi's articles, with one omission: "On Indeterminate Probabilities," from 1974, is one of his most influential contributions to decision theory.

Levi emphasizes that scientific inquiry is a cognitive enterprise. He summarizes his epistemological program as follows:

> The main preoccupation of students of the growth of knowledge should be to explore criteria for evaluating revisions of cognitive states of rational agents (87).

Levi's reason for this is in line with the pragmatist tradition:

> . . . cognitive states provide guidance in deliberation and inquiry. It matters, therefore, how we revise them (90).

On the basis of these general epistemological goals, Levi has, over a number of years, developed a theory of how rational agents ought to revise their cognitive states. This theory is a pathbreaking contribution. Here I want only to raise a few questions concerning some details in his enterprise.

The two main components of Levi's model of a cognitive state are (i) a (potential) corpus $K_{X,t}$ adopted by an agent X at a time t, which consists of all sentences that X accepts as certainly true at t; and (ii) a credal state $B_{X,t}$ which is X's system of personal probability judgments at t. In recent years, Levi has modeled the credal state $B_{X,t}$ by a (convex) set of probability functions Q defined over the given language. Using a set of probability functions, rather than only one as in traditional Bayesianism, allows Levi to represent indeterminate probabilities. This is clearly an improvement over Bayesianism, and Levi exploits this feature in a very interesting way in his decision theory. The connection between the corpus $K_{X,t}$ and the credal state

0022-362X/87/8412/0747$02.00 © 1987 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

$B_{X,t}$ is provided by a "confirmational commitment" that is a function from $K_{X,t}$ to $B_{X,t}$. Levi argues that the role of inductive logic is to impose:

> . . . necessary conditions that a Q-function must satisfy in order to be a member of a potential credal state relative to a potential corpus of knowledge (202).

Thus, in Levi's model, given a confirmational commitment, the credal state $B_{X,t}$ is uniquely determined by the corpus $K_{X,t}$. This is certainly in the tradition of Rudolf Carnap's research program (203), but it has the drawback that $B_{X,t}$ cannot change unless $K_{X,t}$ changes, assuming that the confirmational commitment is kept constant. This seems to be a considerable limitation of Levi's model of cognitive states. In effect, his model excludes changes of personal probabilities which do not lead to changes of the corpus $K_{X,t}$ and thus it unnecessarily limits the class of cognitive inputs that can be accommodated by the agent: for example, the kind of changes covered by so called "Jeffrey conditionalization," i.e., where the probability of a statement h is changed from a value x to a new value y, where y is not identical with 1, cannot be described in a simple way in Levi's model.

To be sure, Levi does not claim that an agent's confirmational commitment cannot be revised, but "[r]ational men should not alter their confirmational commitments for revising credal states without good reason" (204). Thus, Levi views such a change as something exceptional. In the present volume, he does not discuss the conditions under which the agent should or should not revise his commitment. However, in chapter 13 of *The Enterprise of Knowledge,* he makes a tentative proposal concerning some such changes.

It seems much more natural to define a confirmational commitment as a function *from* credal states *to* corpora. If this strategy is adopted, a credal state determines a unique corresponding corpus. In this way, Jeffrey conditionalization is easily accommodated, since a change in the credal state need not involve any change in the corresponding corpus. The simplest way to define such a confirmational commitment would be to say that a sentence h is in $K_{X,t}$ if and only if $Q(h) = 1$ for all Q-functions in $B_{X,t}$. However, it is clear that Levi would not accept this definition. He accepts that, if h is in $K_{X,t}$, then $Q(h) = 1$, for all Q-functions in $B_{X,t}$, but h may, according to Levi, have maximal probability in this sense while the negation of h is still a "serious possibility" in $K_{X,t}$ and thus h is not in $K_{X,t}$ (148). It is not clear to me what further properties h must have in order to be accepted in $K_{X,t}$. In my opinion, defining a confirmational commitment as a function from $B_{X,t}$ to $K_{X,t}$ seems to be the best starting point

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

for specifying what other conditions, apart from having maximal probability, a sentence h must satisfy in order to be accepted in $K_{X,t}$.

Turning next to Levi's theory of the kinematics of cognitive states, we see that he distinguishes between four kinds of changes:

The basic problem, then, is to formulate rationality criteria for expansions, contractions, and replacements. Levi reduces this set of problems in a rather ingenious way by requiring that replacements be describable as sequences of contractions and expansions. Thus, K_2 in (iii) above can be obtained from K_1 by first contracting K_1 with respect to h and then expanding the resulting state by $-h$. And then the central requirement on a contraction with respect to a sentence h is that the loss of informational value be minimized (subject to the constraint that h is given a hearing).

Levi distinguishes between two kinds of expansions—*routine* and *inferential* (90, 116–119). Routine expansion "is the result of implementing a program, to which the agent is committed beforehand, for using the testimony of the senses or of other agents as inputs where the outputs are expansions" (90). Inferential expansion occurs when an agent X is concerned to answer some system of questions. "For each such question he has identified a system of potential answers. A potential answer is an expansion strategy—i.e., a way of adding new information to X's corpus which will somehow meet the demands of information implicit in the question" (116).

The notion of routine expansion seems somewhat problematic, at least in the case when the agent is human. Levi admits that the process can produce inconsistencies in the cognitive state: "Thus, if in response to sensory stimulation, X makes an observation report inconsistent with what is already in his corpus and, hence, certainly false from the point of view he adopted prior to carrying out the observation routine, he is committed to carry through and add the report to his corpus. This is so even if the result is to convert his corpus into an inconsistent one" (119).

Machines that handle information may use routine expansions, but it seems doubtful that such a process is ever used by human agents. A human agent never performs a pure routine expansion, but always "tests" the input in some way or other so that contradictions, at least the most immediate ones, are avoided. Sometimes this may, of course, lead to considerable revisions of the agent's corpus, but this is the price that must be paid by a rational agent of the kind Levi considers for avoiding inconsistencies. (In *The Enterprise of*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Knowledge, pp. 10/1, he argues that, ideally, the cognitive states of rational agents should be equilibria. In my opinion, however, an inconsistent state cannot be an equilibrium point.) And, in this case, the resulting change is not an expansion but a replacement.

When a contraction of a cognitive state takes place, the shift is made from a corpus K_1 to a weaker K_2, i.e., where K_2 is a subset of the set of sentences accepted in K_1. Levi requires that this change be mirrored on the "credal" states B_1 and B_2 by a process he calls "inverse temporal credal conditionalization" (149). This means that, if K_2 is obtained from K_1 by giving up belief in a sentence e (and all sentences that imply e), then for all probability functions Q' in B_2, there is some probability function Q in B_1, and for every Q in B_1 there is a Q' in B_2 such that $Q(h/f) = Q'(h/f\&e)$ for all h in the language and for all f consistent with K_1.

Now, it is routine to verify that, if h is in K_1 (and thus $Q(h) = 1$ for all Q in B_1) and K_2 comes from K_1 by contraction with respect to e, then the sentence "$e \supset h$" will have maximal probability in B_2, i.e., $Q'(e \supset h) = 1$ for all Q' in B_2, if B_2 comes from B_1 by inverse temporal credal conditionalization.[1] According to Levi (148), this alone does not entail that "$e \supset h$" is in K_2, but in most practical cases it will be. If we assume that it is, for all h, then it follows that, if we expand K_2 by e, we will get precisely K_1 back again. I have defended this "recovery" property of contractions elsewhere, and, in my opinion, this is an advantage of Levi's model of contractions. Furthermore, it fits well with his requirement that a contraction involve a minimal loss of information. However, Levi has contested the validity of the recovery property. The simple result above shows that it will be difficult for him to avoid the conclusion that his own model satisfies the recovery postulate.

Finally, a comment about Levi's decision theory. In traditional Bayesian theory, the credal state of an agent is represented by a single probability function which is then used for calculating the expected utilities of the different choice alternatives. Levi generalizes this by modeling a credal state B by a set of probability functions. I think this model has decisive advantages over the traditional Bayesian model. In particular, it can be used to handle the vast area of decision making that lies between the extremes that are commonly called "decision making under risk" and "decision making under uncertainty" (228).

[1] Let Q' be any probability function in B_2. Then $Q'(e \supset h) = Q'(-e \vee h) = Q'(-e) + Q'(h) - Q'(-e\&h) = Q(-e/e) + Q(h/e) - Q(-e\&h/e) = Q(h/e)$, for some Q in B_1. But, since $Q(h) = 1$, it follows that $Q(h/e) = 1$.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The question for a decision theory, then, is how the set B of probability functions should be exploited in decision making. In the present volume, as well as in *The Enterprise of Knowledge,* Levi requires that the first step in forming a decision be to pick out those alternatives which are assigned maximum expected utility for some probability function Q in B. Levi calls these alternatives "E-admissible." He then proceeds to formulate additional restrictions (P-admissibility and S-admissibility) on the alternatives which further delimit the set of optimal alternatives (in most cases there will only be one S-admissible alternative).

Since E-admissibility plays such a central role in Levi's decision theory, it is somewhat surprising to note that he hardly supplies any arguments for this principle. Apparently, the idea is to retain some of the motivation behind maximizing expected utility as used in traditional Bayesian theory. Levi says that other decision criteria "should be invoked only after considerations of expected utility have failed to render a verdict. In resisting the dogmatism of Bayesianism, we should not ignore its insights" (233). But, there are many other ways of generalizing the Bayesian principle of maximizing expected utility, and Levi never shows why E-admissibility is the best one.

On the contrary, I believe that E-admissibility sometimes rejects the wrong alternatives. The following modification of one of Levi's examples (*The Enterprise of Knowledge,* pp. 157/8) can be used for my purposes. Let h be the statement that the next toss of coin a will land heads. Let Q_1 be a probability function such that $Q_1(h) = .4$ and Q_2 a probability function such that $Q_2(h) = .6$. Let B consist of Q_1, Q_2 and all convex combinations of these functions, i.e., all functions Q such that, for all e, $Q(e) = r \cdot Q_1(e) + (1 - r) \cdot Q_2(e)$, for some r, $0 < r < 1$. Let G_1 be a gamble such that, if a lands heads you win \$1000, otherwise you loose \$995. G_2 is the gamble where you lose \$995 if a lands heads, and where you win \$1000 otherwise. G_3 is the option of refusing to gamble, with neither loss nor gain. According to Levi's theory, only G_1 and G_2 are E-admissible. However, I conjecture that most people would choose G_3 in this decision situation. Furthermore, this slightly risk-averse behavior seems perfectly rational. Levi's requirement of E-admissibility has the consequence that, in many cases, the choices recommended by his theory seem unrealistically risk prone. At least one would like to see some further arguments for why a decision maker should be required to choose among the E-admissible alternatives.

In general, I find Levi's approach to the philosophical problems very fruitful. The critical points I have raised are mainly marginal comments on his epistemological program of characterizing rational

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

revisions of knowledge and belief and on his decision theory which generalizes the traditional Bayesian theory. However, the book contains stimulating discussions on a number of other areas which I have not been able to touch upon here.

PETER GÄRDENFORS

Australian National University and Lund University

Der philosophische Diskurs der Moderne. Zwölf Vorlesungen. JÜRGEN HABERMAS. Frankfurt: Suhrkamp, 1985. 450 p. *The Philosophical Discourse of Modernity: Twelve Lectures,* Frederick G. Lawrence, trans. Cambridge, Mass.: MIT Press, 1987. 422 p. $27.50.

To appreciate the set of problems dealt with in Jürgen Habermas's *Der philosophische Diskurs der Moderne,* the reader must keep in mind two background experiences that have deeply colored the European, in contrast to the North American, experience of modernity. The first is the memory of an old order, of an organic community preceding commercialization, capitalism, urbanization, and industrialism. The second is the emergence of mass fascist movements in the twentieth century with their troublesome combination of antimodernist social nostalgia and technical ultramodernism, their mixing of aesthetics and politics, and their vacillation between ruthless individualism and mindless collectivism. Whereas the memory of a lost organic community stands at the beginning of the confrontation with modernity and is an attempt to limit it via a return to the past, fascist movements—and in particular National Socialism—indicate the exhaustion of the modern project and express the inhumanity to which a modern consciousness, unable to limit itself normatively, can ultimately lead.

It is against the background of such a memory, based partly on fiction and partly on fact, that Hegel described the new social order inaugurated by the modern period as *Entzweiung,* as the division or sundering apart of individual and community, self-interest and the common good, morality and politics, or, in metaphysical terms, of particularity and universality. The social order that had not experienced such bifurcation was named *Sittlichkeit,* an ethical community. After his early love affair with the ancient Greek *polis,* which he idealized as the paradigm of *Sittlichkeit,* Hegel settled down to a more sober relationship with modernity while still being haunted by

0022-362X/87/8412/0752$02.00 © 1987 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the questions: How could *Entzweiung* be overcome? How could the divisions that afflict the modern world be healed even when one recognized as valid the achievements of modernity? "Hegel," Habermas writes, "is not the first philosopher who belongs to modern times, but he is the first for whom modernity has become a problem. First in his theory does the conceptual relation between modernity, time consciousness, and rationality become visible. At the end Hegel himself explodes apart this constellation, for a rationality which has puffed itself up as Absolute Spirit neutralizes the very conditions under which the moderns have attained a consciousness of themselves" (57).[1]

As the debate on Hegel's legacy indicates, beginning with the division of his disciples into the Left and the Right Hegelians, Hegel's synthesis of rationality and actuality, the reconciliation of modern consciousness with modern times, has failed. Hegel made the totality seem rational at the cost of rationalizing the irrational, the merely contingent, the unjust, and the idiosyncratic. Thus, the dilemma which dominates modern consciousness and which Hegel clearly saw but could not solve returns: "A modernity that is freed of images of authority, that is open to the future, and addicted to continuous self-renewal can only create itself the criteria to judge itself by. Thus, the principle of subjectivity emerges as the only origin of normativity which grows out of the time consciousness of modernity itself" (55). But can this principle of subjectivity alone and "the structure of self-consciousness intrinsic to it be adequate sources of normative orientation . . . can it stabilize a historical formation which has freed itself of all historical ties?" (31).

Habermas interprets the philosophical discourse of the moderns as so many attempts to answer this question. Whereas Hegel had distinguished between Enlightenment reason, *Verstand,* and true rationality, *Vernunft,* the discourse of the moderns since the days of the Young Hegelians, condemns a "reason which is grounded in the principle of subjectivity" (70). The radical critique of reason, started by the Young Hegelians, is given its sharpest formulation by Nietzsche.

Nietzsche initiates two types of critiques of reason. First, there is the attitude of the *skeptical naturalist* who sees reason as an instrument of adaptation of the species, as an organ developed by animals too weak to survive without a lie (remember Nietzsche's definition of truth as "a lie invented by a species who could not survive otherwise"). Such reason is the expression of the will to power. This

[1] All translations in this review are mine.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

attitude of the skeptical naturalist, who dissects the physiological origins of reason, is continued in the twentieth century by Georges Bataille and Michel Foucault. The second kind of critique of reason defends an *artistic* or *poetic* conception of the world. An obsolete Western metaphysics is now replaced by art and poetry. They affirm life and recount the heights to which an unbounded subjectivity in search of ever new experiences of transcendence can soar. They also document the pits of despair into which the modern individual can fall once its meaning-creating powers have dried up. Martin Heidegger and Jacques Derrida, Habermas claims, follow Nietzsche, the artist, and the poet. They push Western metaphysics to those extremes where it confronts the limits of discursive thought and transforms itself into another kind of discourse—poetry or deconstruction (120 ff.).

Marx and Nietzsche, as well as Bataille, Derrida, and Foucault, put modern reason on trial for merely denouncing repression and exploitation, humiliation and alienation, only to replace them by the "irreproachable domination of rationality itself" (70). The experience of modernity is characterized by a dialectic of Enlightenment: Enlightenment reason is itself the creator of the domination that it ostensibly denounces. If this is so, however, the dilemma of modernity is even more acute. If the principle of subjectivity is the only source of normative orientation for the moderns but the realization of this principle implies a dialectic of Enlightenment, can one provide a rational normative orientation for modern consciousness at all?

Habermas argues that, for Nietzsche and Heidegger, as well as for Bataille, Foucault, and Derrida (and even for Niklas Luhmann), the answer is negative. He also maintains that this answer could not but be negative, since all these thinkers have failed to overcome the philosophy of consciousness or of the subject, and have thus failed to explore the real normative potential of modernity. The philosophy of consciousness or of the subject privileges the relationship of the thinking or producing individual to an external world, and takes as its philosophical beginning point the experience of "reflection" of the individual subject. On the surface, this claim is highly implausible. Nietzsche as well as Heidegger, Derrida as well as Foucault, would themselves appear as critics of the philosophy of consciousness or of the subject. Through a detailed analysis of the work of these authors, full of surprising and fascinating excursions—like Heidegger's "turn" (*Kehre*) in the context of his national-socialist sympathies, Derrida's theory of language and Jewish mysticism, and the empirical problems of Foucault's theory of modernity—Ha-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

bermas repeatedly emphasizes that none of these thinkers take as their starting point the experience of human subjects communicating linguistically, situated in a lifeworld, and for whom the claims of reason are embedded in structures of everyday communication itself (368 ff.).

Der philosophische Diskurs der Moderne is thus Habermas's answer to his poststructuralist and postmodernist critics. These critics, like Jean-François Lyotard, have seen in Habermas's defense of the ideals of the Enlightenment—rationality, autonomy, consensus—just one more "grand narrative" to rescue a discredited Western rationality. Habermas's answer to this charge is to show why the radical critique of reason leads to a series of aporias, self-referential paradoxes, and absurdities. The fundamental paradox, common to all radical critics of reason, is that, as thinkers and philosophers, they must use the very same reason in their discourse which they wish to unmask, denounce, and accuse.

Now, among the discourses of modernity considered by Habermas, there is one in particular which is aware of this criticism and which preempts it by incorporating it into its very content. This is the discourse of the deconstructionist. The deconstructionist knows that s/he will be accused of self-referential absurdities, of aporetic formulations, of self-defeating arguments by those who still take seriously the claim of reason to justify itself. To escape the game of justification, to escape defeating reason by rational means, the deconstructionist self-consciously creates a language that places paradox, aporia, and absurdity at its center. The deconstructionist has seen the rationalist looking at her/him, and has made this glance its own, while all the same continuing to observe the rationalist as a member of a deranged species. Derrida is not the only master of such a self-consciously paradoxical discourse. Theodor Adorno's critique of identity logic already anticipates the deconstructionist project.

The discourse of the deconstructionist poses a special challenge to Habermas: In what ways is the charge of self-referential absurdities damaging for a mode of thinking that only too willingly acknowledges this to be its virtue? What if the deconstructionist merely chooses to play with the rationalist? What then? I can hardly do justice within the scope of this review to Habermas's attempt to defeat Derrida at his own game. Through a compelling reading of Derrida's Husserl critique, and through an extremely interesting analysis of Derrida's philosophy of language in the context of the Jewish mystical tradition, Habermas focuses the confrontation with the deconstructionist around the issue: Which conception of language is more fundamental—the view of language as *rhetoric,* or the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pragmatic view of language as *communication* which is always implicitly oriented toward the validity claims of truth, truthfulness, and normative rightness?

In my opinion, *Der philosophische Diskurs der Moderne* does not resolve this question. Indeed, it is unclear what a resolution to this issue might look like given the postmetaphysical and nonfoundationalist position that Habermas accepts. How could one argue that one conception of language was more "fundamental" than another without begging the question and without already presupposing the validity of what one sets out to prove? This question is crucial, for Habermas's confrontations with Derrida and Heidegger (and with Cornelius Castoriadis) turn on differing conceptions of language. In addition to the Derridean view of language as rhetoric, there is also the Heideggerian view, subscribed to by Castoriadis (369 ff.), of language as "world-disclosing" (*welterschließend*). But Habermas can trace the failure of the various discourses of modernity to adherence to the philosophy of the subject only if he can also succeed in arguing, without begging the question, for the priority of the *pragmatic* conception of language, centered around speech acts as constituent units, over other accounts of language. Of course, one could always claim, as Richard Rorty[2] has done recently, that no "argument," in the traditional sense of the term as a chain of reasoning demanding assent or dissent, can be provided at this point. One can at most present "metaphors" and metaphorical narratives of how self and community would have to be thought of once a particular view of language were adapted.

In face of this challenge, the only argumentative strategy open to Habermas which I can see is to reformulate what "being more fundamental than" or "having priority over" can mean. Such an argument demonstrating priority or fundamentality cannot be thought of as a deductive chain of reasoning. Habermas has elsewhere characterized his method of preferred philosophical proof as a kind of "coherence test."[3] According to this procedure, a postmetaphysical and postfoundationalist philosophy can argue for the validity of its premises only by showing through a kind of "reflective equilibrium" (A) that these premises lead to fruitful empirical research in the human and social sciences; and (B) that these sciences, since they are themselves dependent upon philosophical assumptions, cannot verify or falsify them but can provide evidence for their plausibility or

[2] Cf. Richard Rorty, "The Contingency of Language," *London Review of Books*, 17 April 1986; "The Contingency of Selfhood," *London Review of Books*, 8 May 1986; and "The Contingency of Community," *London Review of Books*, 24 July 1986.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

implausibility. As is well-known, the human sciences that Habermas emphasizes are the genetic epistemology of Jean Piaget and the cognitive, developmental psychology of Lawrence Kohlberg.

This strategy has many problems—some of which have to do with Habermas's interpretation of the place of these empirical theories within their own research domains. Other questions concern the strategy of indirect philosophical proof itself: Can this method avoid either falling into circularity or into some kind of scientism? Obviously these issues cannot be settled here. Suffice it to say that this strategy of "indirect proof" expands the debate with the postmodernist in an interesting way. Between foundationalism, on the one hand, and deconstructionism, on the other, may lie a third option for philosophy at the present which requires that it abandon its position at the bottom or at the top of the hierarchy of knowledge and enter into serious collaboration with the human and social sciences. Can such a postmetaphysical and postfoundationalist conception of philosophy succeed where Hegel and all the others have failed, namely, in forfeiting philosophy's claim of absolute certainty while insisting on the legacy of the Enlightenment? In other words, is a universalistic, nonfoundationalist philosophy possible?

Der Philosophische Diskurs der Moderne poses this question in all its urgency. In taking the critics of the Enlightenment like Heidegger, Bataille, Foucault, and Derrida so seriously, this work succeeds in being a hermeneutical *tour de force* that truly excels in the art of immanent reading and analysis. Habermas does not have the last word in the current debate with poststructuralists like Lyotard and antifoundationalists like Rorty, nor is it desirable for that matter that anyone should have the last word. So far, Rorty is right: To keep the conversation going is what we can hope for. With this work, however, Habermas has succeeded in throwing the ball into his critics' court. It is to be wished that they will throw it back with the same elegance and rigor.

SEYLA BENHABIB

Harvard University

[3] Cf. Habermas, *The Theory of Communicative Action*, vol. 1, Thomas A. McCarthy, trans. (Boston: Beacon Press, 1984), pp. 2/3; and Habermas's answer to Rorty, "Die Philosophie als Platzhalter und Interpret," in *Moralbewußtsein und kommunikatives Handeln* (Frankfurt: Suhrkamp, 1983), pp. 9–29 [English trans. in Ken Baynes, James Bohman, and Thomas A. McCarthy, eds., *After Philosophy* (Cambridge, Mass.: MIT Press, 1986), pp. 296–315.]

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NAMES OF CONTRIBUTORS ARE PRINTED IN SMALL CAPITALS

ADDELSON, KATHRYN PYNE: Autonomy and Respect [abstract]. 628 (11).
ADLER, JONATHAN E.: Relevant Alternatives, Presuppositions, and Skepticism [abstract]. 653 (11).
ALEXANDER, LARRY A.: Scheffler on the Independence of Agent-centered Prerogatives from Agent-centered Restrictions. 277 (5).
American Philosophical Association, Eastern Division, 84th Annual Meeting. Symposium Papers, 521 (10), 619 (11); Program 585 (10).
Announcements. 344 (6), 398 (7).
Aristotle (Hylomorphism in—). CHARLOTTE WITT. 673 (11).
Aristotle's *Metaphysics*. APA Symposium. 666 (11).
Aristotle's *Metaphysics* [abstract]. SARAH BROADIE. 679 (11).
Aristotle's Solution to the Problem of Sensible Substance. EDWARD HALPER. 666 (11).
Atemporal Duration: A Reply to Fitzgerald. ELEONORE STUMP and NORMAN KRETZMANN. 214 (4).
Authors (Memo to—). 56 (1).
Autonomy and Respect [abstract]. KATHRYN PYNE ADDELSON. 628 (11).

BAKER, LYNNE RUDDER: Content by Courtesy. 197 (4).
BARBER, RICHARD: Public Policy and the Allocation of Scarce Medical Resources. 655 (11).
BARON, MARCIA: Kantian Ethics and Supererogation. 237 (5).
BATKIN, NORTON: Review of Richard Wollheim: *The Thread of Life*. 336 (6).
BEALER, GEORGE: The Boundary Between Philosophy and Cognitive Science [abstract]. 553 (10).
BELNAP, NUEL, and ANIL GUPTA: A Note on Extension, Intension, and Truth. 168 (3).
BENHABIB, SEYLA: Review of Jürgen Habermas: *Der philosophische Diskurs der Moderne: Zwölf Vorlesungen*. 752 (12).
BENSON, PAUL: Freedom and Value. 465 (9).
BIGELOW, JOHN, and ROBERT PARGETTER: Functions. 181 (4).
BITTNER, RÜDIGER: Review of M. J. Inwood's *Hegel*. 512 (9).
BLAIS, MICHEL J.: Epistemic Tit for Tat. 363 (7).
Blumenberg, Hans: *Lebenszeit und Weltzeit*. Review by KARSTEN HARRIES. 516 (9).
Boundary Between Philosophy and Cognitive Science (The—) [abstract]. GEORGE BEALER. 553 (10).
Brand, Myles: *Intending and Acting: Toward a Naturalized Action Theory*. Review by PETER SLEZAK. 49 (1).
BRANDON, ROBERT N.: Review of Alexander Rosenberg: *The Structure of Biological Science*. 224 (4).
BROADIE, SARAH: Aristotle's *Metaphysics* [abstract]. 679 (11).
BROOME, JOHN: Utilitarianism and Expected Utility. 405 (8).
BUDD, MALCOLM: Review of Anthony Kenny: *The Legacy of Wittgenstein*. 42 (1).

Causality (Conceptions of—in Early Modern Philosophy). APA Symposium. 567 (10).
CHURCHLAND, PATRICIA SMITH: Epistemology in the Age of Neuroscience. 544 (10).
Cognitive Science and Metaphysics. ALVIN I. GOLDMAN. 537 (10).
Compatibilism and the Argument from Unavoidability. THOMAS P. FLINT. 423 (8).
Conceptions of Causality in early Modern Philosophy. APA Symposium. 567 (10).
Content by Courtesy. LYNNE RUDDER BAKER. 197 (4).
COOK, J. THOMAS: Deciding to Believe without Self-Deception. 441 (8).
Copyright (Memo on—). 119 (2).
Cultural Divergence: A Wittgensteinian Paradox (Explaining—). PAUL SEABRIGHT. 11 (1).

DALLMAYR, FRED: The Discourse of Modernity: Hegel and Habermas. 681 (11).
DANIELS, NORMAN: Justice and the Dissemination of "Big-ticket" Technologies [abstract]. 664 (11).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Dead (The—). PALLE YOURGRAU. 84 (2).
Deciding to Believe without Self-Deception. J. THOMAS COOK. 441 (8).
De NYS, MARTIN J.: Political Representation and Economic Liberty [abstract]. 565 (10).
Descartes (How God Causes Motion:—, Divine Sustenance, and Occasionalism. DANIEL GARBER. 567 (10).
Dilemmas (Moral Realisms and Moral—). WALTER SINNOTT-ARMSTRONG. 263 (5).
The Discourse of Modernity: Hegel and Habermas. FRED DALLMAYR. 681 (11).
Divine Impulse, Voluntary Motion, and Intelligible Causal Models [abstract]. EILEEN O'NEILL. 580 (10).
Does "Fitness" Fit the Facts? A Reply to Williams and Rosenberg's Rejoinder. ELLIOTT SOBER. 220 (4).
DRYDYK, JAY: Liberties, Wrongs, and Representation. 556 (10).
Duration: A Reply to Fitzgerald (Atemporal—). ELEONORE STUMP and NORMAN KRETZMANN. 214 (4).
Dworkin, Ronald: *A Matter of Principle* and *Law's Empire*. Review by KENT GREENAWALT. 284 (5).

ELGIN, CATHERINE Z., and ISRAEL SCHEFFLER: Mainsprings of Metaphor. 331 (6).
Enlightenment and Rationality. ALEX HONNETH. 692 (11).
Enlightenment and Rationality. APA Symposium. 681 (11).
Enlightenment and Reason [abstract]. TOM ROCKMORE. 698 (11).
Epistemic Tit for Tat. MICHEL J. BLAIS. 363 (7).
Epistemology and Philosophy of Mind. APA Symposium. 537 (10).
Epistemology in the Age of Neuroscience. PATRICIA SMITH CHURCHLAND. 544 (10).
Epistemology of First-person Reference (The—). CAROL A. ROVANE. 147 (3).
Essence, and Indiscernibility (Identity, —). STEPHEN YABLO. 293 (6).
Ethics and Supererogation (Kantian—). MARCIA BARON. 237 (5).
Expected Utility (Utilitarianism and—). JOHN BROOME. 405 (8).
Explaining Cultural Divergence: A Wittgensteinian Paradox. PAUL SEABRIGHT. 11 (1).
Extension, Intension, and Truth (A Note on—). ANIL GUPTA and NUEL BELNAP. 168 (3).

Feminist Perspectives on Individual Choice. APA Symposium. 619 (11).
First Look at the Pornography/Civil Rights Ordinance: Could Pornography Be the Subordination of Women? (A—). MELINDA VADAS. 487 (9).
First-person Reference (The Epistemology of—). CAROL A. ROVANE. 147 (3).
First Person (Indeterminacy, Empiricism and the—). JOHN R. SEARLE. 123 (3).
"Fitness" Fit the Facts? A Reply to Williams and Rosenberg's Rejoinder (Does—). ELLIOTT SOBER. 220 (4).
Fitzgerald (Atemporal Duration: A Reply to—). ELEONORE STUMP and NORMAN KRETZMANN. 214 (4).
FLINT, THOMAS P.: Compatibilism and the Argument from Unavoidability. 423 (8).
FODOR, J. A.: Making Mind Matter More [abstract]. 642 (11).
Freedom and Value. PAUL BENSON. 465 (9).
Frege and Husserl on Sense [abstract]. Donn Welton. 535 (10).
Frege (Husserl and—). APA Symposium. 521 (10).
Functions. JOHN BIGELOW and ROBERT PARGETTER. 181 (4).

GARBER, DANIEL: How God Causes Motion: Descartes, Divine Sustenance, and Occasionalism. 567 (10).
GÄRDENFORS, PETER: Review of Isaac Levi: *Decisions and Revisions: Philosophical Essays on Knowledge and Value*. 747 (12).
GOLDMAN, ALAN H.: Red and Right. 349 (7).
GOLDMAN, ALVIN I.: Cognitive Science and Metaphysics. 537 (10).
GREENAWALT, KENT: Review of Ronald Dworkin: *A Matter of Principle* and *Law's Empire*. 284 (5).
GUPTA, ANIL, and NUEL BELNAP: A Note on Extension, Intension, and Truth. 168 (3).

HABERMAS, JÜRGEN: *Der philosophische Diskurs der Moderne: Zwölf Vorlesungen*. Review by Seyla Benhabib. 752 (12).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Habermas. (The Discourse of Modernity: Hegel and—). FRED DALLMAYR. 681 (11).
HALPER, EDWARD: Aristotle's Solution to the Problem of Sensible Substance. 666 (11).
Hansen, Chad: *Language and Logic in Ancient China*. Review by MICHAEL R. MARTIN. 37 (1).
Harmon, Gilbert: *Change in View: Principles of Reasoning*. Review by ISAAC LEVI. 376 (7).
HARRIES, KARSTEN: Review of Hans Blumenberg's *Lebenszeit und Weltzeit*. 516 (9).
Hegel (The Discourse of Modernity:—and Habermas). FRED DALLMAYR. 681 (11).
HONNETH, ALEX: Enlightenment and Rationality. 692 (11).
Hornstein, Norbert: *Logic as Grammar*. Review by SCOTT SOAMES. 447 (8).
How God Causes Motion: Descartes, Divine Sustenance, and Occasionalism. DANIEL GARBER 567 (10).
Human Beings. MARK JOHNSTON. 59 (2).
HURKA, THOMAS: The Well-rounded Life. 727 (12).
Husserl and Frege. APA Symposium. 521 (10).
Husserl and Frege: ROBERT SOKOLOWSKI. 521 (10).
Husserl and Frege. RONALD McINTYRE. 528 (10).
Hylomorphism in Aristotle. CHARLOTTE WITT. 673 (11).

Identity, Essence, and Indiscernibility. STEPHEN YABLO. 293 (6).
Indeterminacy, Empiricism, and the First Person. JOHN R. SEARLE. 123 (3).
Indeterminacy of Translation Again. W. V. QUINE. 5 (1).
Indiscernibility (Identity, Essence, and—). STEPHEN YABLO. 293 (6).
Individual Choice (Feminist Perspectives on—). APA Symposium. 619 (11).
INWOOD, M. J.: *Hegel*. Review by RÜDIGER BITTNER. 512 (9).

JOHNSEN, BREDO C.: Relevant Alternatives and Demon Skepticism. 643 (11).
Johnsonian Prize in Philosophy. 57 (1).
JOHNSTON, MARK: Human Beings. 59 (2).
Justice and the Dissemination of "Big-Ticket" Technologies. NORMAN DANIELS. 664 (11).

Kantian Ethics and Supererogation. MARCIA BARON. 237 (5).
Kenny, Anthony: *The Legacy of Wittgenstein*. Review by MALCOLM BUDD. 42 (1).
KITCHER, PHILIP, and WESLEY SALMON: Van Fraassen on Explanation. 315 (6).
KRETZMANN, NORMAN, and ELEONORE STUMP: Atemporal Duration: A Reply to Fitzgerald. 214 (4).
KYBURG, HENRY E., JR.: Review of Patrick Suppes: *Probabilistic Metaphysics*. 45 (1).

LARMORE, CHARLES: Review of Barry Stroud: *The Significance of Philosophical Scepticism*. 384 (7).
Le PORE, ERNEST, and BARRY LOEWER: Mind Matters. 630 (11).
LEVI, ISAAC: *Decisions and Revisions: Philosophical Essays on Knowledge and Value*. Review by Peter Gärdenfors. 747 (12).
LEVI, ISAAC: Review of Gilbert Harman: *Change in View: Principles of Reasoning*. 376 (7).
Liberties, Wrongs, and Representation. APA Symposium. 556 (10).
Liberties, Wrongs, and Representation. JAY DRYDYK. 556 (10).
LOEWER, BARRY, and ERNEST Le PORE: Mind Matters. 630 (11).

MacLane, Saunders: *Mathematics: Form and Function*. Review by COLIN McLARTY. 33 (1).
McGINN, COLIN: Review of Sydney Shoemaker: *Identity, Cause, and Mind*. 227 (4).
McINTYRE, RONALD: Husserl and Frege. 528 (10).
McLARTY, COLIN: Review of Saunders MacLane: *Mathematics: Form and Function*. 33 (1).
Mainsprings of Metaphor. CATHERINE Z. ELGIN and ISRAEL SCHEFFLER. 331 (6).
Making Mind Matter More [abstract]. J. A. FODOR. 642 (11).
MARTIN, MICHAEL R.: Review of Chad Hansen: *Language and Logic in Ancient China*. 37 (1).
Medical Resources (Public Policy and the Allocation of Scarce—). APA Symposium. 655 (11).
Memo on Copyright. 119 (2).
Memo to Authors. 56 (1).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Metaphor (Mainsprings of—). CATHERINE Z. ELGIN and ISRAEL SCHEFFLER. 331 (6).
MEYERS, DIANA T.: Personal Autonomy and the Paradox of Feminine Socialization. 619 (11).
Mind (Epistemology and Philosophy of—). APA Symposium. 537 (10).
Mind Matters. APA Symposium. 630 (11).
Mind Matters. ERNEST Le PORE and BARRY LOEWER. 630 (11).
Modern Philosophy (Conceptions of Causality in Early—. APA Symposium. 567 (10).
Modernity (The Discourse of—: Hegel and Habermas). FRED DALLMAYR. 681 (11).
Moral Realisms and Moral Dilemmas. WALTER SINNOTT-ARMSTRONG. 263 (5).
Moral Theory and Moral Alienation. ADRIAN M. S. PIPER. 102 (2).

New Annuals and Proceedings. 176 (3).
New Anthologies in Paperback. 177 (3).
New Books. 175 (3), 234 (4), 398 (7), 456 (8).
New Books in Paperback. 178 (3).
New Editions and Translations. 175 (3).
Note on Extension, Intension, and Truth (A—). ANIL GUPTA and NUEL BELNAP. 168 (3).
Notes and News. 54 (1), 120 (2), 233 (4), 292 (5), 345 (6), 520 (9), 582 (10), 702 (11).

O'NEILL, EILEEN: Divine Impulse, Voluntary Motion, and Intelligible Causal Models [abstract]. 580 (10).

Paradox of Feminine Socialization (Personal Autonomy and The—). DIANA T. MEYERS. 619 (11).
PARGETTER, ROBERT, and JOHN BIGELOW: Functions. 181 (4).
Personal Autonomy and the Paradox of Feminine Socialization. DIANA T. MEYERS. 619 (11).
Philosophical Letter of Alfred Tarski (A—). MORTON WHITE. 28 (1).
PIPER, ADRIAN M. S.: Moral Theory and Moral Alienation. 102 (2).
Political Representation and Economic Liberty [abstract]. MARTIN J. De NYS. 565 (10).
Pornography/Civil Rights (A First Look at the— Ordinance: Could Pornography Be the Subordination of Women?). 487 (9).
Prerogatives from Agent-centered Restrictions (Scheffler on the Independence of Agent-centered—). LARRY A. ALEXANDER. 277 (5).
Prize in Philosophy (Johnsonian—). 57 (1).
Public Policy and the Allocation of Scarce Medical Resources. APA Symposium. 655 (11).
Public Policy and the Allocation of Scarce Medical Resources. RICHARD BARBER. 655 (11).

QUINE, W. V.: Indeterminancy of Translation Again. 5 (1).

Rationality (Enlightenment and—). APA Symposium. 681 (11).
Realisms and Moral Dilemmas (Moral—). WALTER SINNOTT-ARMSTRONG. 263 (5).
Red and Right. ALAN H. GOLDMAN. 349 (7).
Relevant Alternatives and Demon Skepticism. APA Symposium. 643 (11).
Relevant Alternatives and Demon Skepticism. BREDO C. JOHNSEN. 643 (11).
Relevant Alternatives, Presuppositions, and Skepticism [abstract]. JONATHAN E. ADLER. 653 (11).
Representation (Liberties, Wrongs, and—). APA Symposium. 556 (10).
ROCKMORE, TOM: Enlightenment and Reason [abstract]. 698 (11).
Rosenberg, Alexander: *The Structure of Biological Science*. Review by ROBERT N. BRANDON. 224 (4).
Rosenberg's Rejoinder (Does "Fitness" Fit the Facts? A Reply to Williams and—). ELLIOTT SOBER. 220 (4).
Ross, J. F.: *Portraying Analogy*. Review by JOSEF STERN. 392 (7).
ROVANE, CAROL A.: The Epistemology of First-person Reference. 147 (3).

SALMON, WESLEY, and PHILIP KITCHER: Van Fraassen on Explanation. 315 (6).
SCHEFFLER, ISRAEL, and CATHERINE Z. ELGIN: Mainsprings of Metaphor. 331 (6).
Scheffler on the Independence of Agent-centered Prerogatives from Agent-centered Restrictions. LARRY A. ALEXANDER. 277 (5).
SEABRIGHT, PAUL: Explaining Cultural

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Divergence: A Wittgensteinian Paradox. 11 (1).
SEARLE, JOHN R.: Indeterminacy, Empiricism, and the First Person. 123 (3).
Serious Philosophy and Freedom of Spirit. ERNEST SOSA. 707 (12).
Shoemaker, Sydney: *Identity, Cause, and Mind.* Review by COLIN McGINN. 227 (4).
SINNOTT-ARMSTRONG, WALTER: Moral Realisms and Moral Dilemmas. 263 (5).
Skepticism (Relevant Alternatives and Demon—). APA Symposium. 643 (11).
SLEZAK, PETER: Review of Myles Brand: *Intending and Acting: Toward a Naturalized Action Theory.* 49 (1).
SOAMES, SCOTT: Review of Norbert Hornstein's *Logic as Grammar.* 447 (8).
SOBER, ELLIOTT: Does "Fitness" Fit the Facts? A Reply to Williams and Rosenberg's Rejoinder. 220 (4).
SOKOLOWSKI, ROBERT: Husserl and Frege. 521 (10).
SOSA, ERNEST: Serious Philosophy and Freedom of Spirit. 707 (12).
STERN, JOSEF: Review of J. F. Ross's *Portraying Analogy.* 392 (7).
Stroud, Barry: The *Significance of Philosophical Scepticism.* Review by CHARLES LARMORE. 384 (7).
STUMPE, ELEONORE, and NORMAN KRETZMANN: Atemporal Duration: A Reply to Fitzgerald. 214 (4).
Substance (Aristotle's Solution to the Problem of Sensible—). EDWARD HALPER. 666 (11).
Supererogation (Kantian Ethics and—). MARCIA BARON. 237 (5).
Suppes, Patrick: *Probabilistic Metaphysics.* Review by HENRY E. KYBURG, JR. 45 (1).

Tarski (A Philosophical Letter of—). MORTON WHITE. 28 (1).
Translation Again (Indeterminacy of—). W. V. QUINE. 5 (1).

Utilitarianism and Expected Utility. JOHN BROOME. 405 (8).
VADAS, MELINDA: A First Look at the Pornography/Civil Rights Ordinance: Could Pornography Be the Subordination of Women? 487 (9).
Value (Freedom and—). PAUL BENSON. 465 (9).
Van Fraassen on Explanation. PHILIP KITCHER and WESLEY SALMON. 315 (6).

Well-rounded Life (The—). THOMAS HURKA. 727 (12).
WELTON, DONN: Frege and Husserl on Sense [abstract]. 535 (10).
WHITE, MORTON: A Philosophical Letter of Alfred Tarski. 28 (1).
Williams and Rosenberg's Rejoinder (Does "Fitness" Fit the Facts? A Reply to—). ELLIOTT SOBER. 220 (4).
WITT, CHARLOTTE: Hylomorphism in Aristotle. 673 (11).
Wittgensteinian Paradox (Explaining Cultural Divergence:—). PAUL SEABRIGHT. 11 (1).

YABLO, STEPHEN: Identity, Essence, and Indiscernibility. 293 (6).
YOURGRAU, PALLE: The Dead. 84 (2).

पुस्तकालय
110147

ANNOUNCEMENTS

The trustees and editors of *The Journal of Philosophy* announce, with some reluctance, an increase in subscription prices. As of January 1, 1988, these will be as follows:
$40/year for institutions
$25/year for individuals
$20/year for students and retired/unemployed philosophers
$6/year for postage outside U.S.A.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Compiled
1999-2000

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri